Estimation of quarticity based on high frequency data

Master's Thesis

in partial fulfillment of the requirements for the degree of

Master of Economics and Management Science

submitted to

Prof. Dr. Wolfgang K. Härdle
Prof. Dr. Ostap Okhrin

Institute for Statistics and Econometrics
Ladislaus von Bortkiewicz Chair of Statistics
C.A.S.E. - Centre of Applied Statistics and Economics

School of Business and Economics
Humboldt-Universität zu Berlin

by
Ivan Vasylchenko
(535624)

Berlin, June 21, 2012

Abstract

Precise estimation of integrated quarticity is highly important, while this value provides inference about integrated volatility and is a valuable ingredient of jump hypothesis test statistics. Estimation of integrated quarticity based on high frequency data created additional challenges, which led to development of new measures, robust to jumps and microstructure noise.

Different combinations of Multipower Volatility Estimators, Nearest Neighbor Truncation Estimators and Robust Neighborhood Truncation Estimators are analyzed in detail. After their application to real market data, each of the estimators is assessed via set of conducted simulation models.

Special attention is paid to the Robust Neighborhood Truncation Estimators which operate on lower order statistics log-returns, while they were prematurely left out of analysis in previous works. Performed simulations as well as empirical calculations proved additional efficiency and jump robustness of these estimators.

Keywords: asset price, integrated volatility, integrated quarticity, high frequency data, market microstructure noise, jump robustness

Contents

List of Figures

List of Tables

List of Abbreviations

BM	Brownian Motion
BQ	Bipower Quarticity Estimator
BV	Bipower Variation Estimator
DJIA	Dow Jones Industrial Average
HOS	Higher Order Statistics RNT Estimator
IQ	Integrated Quarticity
IV	Integrated Volatility
LOS	Lower Order Statistics RNT Estimator
MPV	Realized Multipower Variation Estimator
NNT	Nearest Neighbor Truncation Estimator
NT	Neighborhood Truncation Estimator
QQ	Quad-power Quarticity
RMSE	Realized Mean Squared Error
RNT	Robust Neighborhood Truncation Estimator
RQ	Realized Quarticity
RV	Realized Volatility
SV-U	Stochastic Volatility Model with U-shape Intraday Pattern
TQ	Tri-power Quarticity Estimator
TV	Tri-power Variation Estimator

Introduction

Last decade financial markets were highlighted with emergence and rapid development of the new industry sector - high frequency trading. Some years ago it took transactions more then ten seconds in order to execute, while nowadays hundreds of them can squeeze in one second. Such a change was mainly driven by decimilization of trading prices and advances in technologies: computational powers and data transfer speeds have grown exponentially. While such operating speeds are unreachable for human trading, more and more market participants started building up computational facilities and developing quantitative algorithms with a goal to outperform competitors.

Eventually, these market transformations have led to generation of enormous amounts of high frequency data sets, which due to their structure sometimes require review of statistical approaches or deduction of radically new ones. Estimation of integrated volatility and integrated quarticity is one of those questions, which have gained a lot of attention in recent years. Irregularity of the intraday returns of the asset price within high frequency data sets coupled with microstructure noise required new robust approaches to estimating these values, thus, extensive work in this direction was conducted by solid number of authors.

Andersen et al. (2001) first introduced the complementary volatility measure, termed realized volatility. Latter is coupled together with realized quarticity measure.

Bipower variation, as an initial term in multipower variation estimator theory, was proposed by Barndorff-Nielsen and Shephard (2004). This paper shows that introduced realized bipower variation dispose some robustness to jumps in price processes. It was demonstrated that realized bipower variation can estimate integrated power volatility in stochastic volatility models and moreover, under some conditions, it can be a good measure to integrated variance in the presence of jumps.

Authors Andersen et al. (2009) came up with two new jump-robust estimators of integrated variance based on high frequency return observations, namely MinRV and MedRV. Their findings prove that these estimators can be good alternative to the multipower variation estimators.

Andersen et al. (2011) presented the family of efficient robust neighborhood truncation (RNT) estimators for the integrated power variation based on the order statistics of a set of unbiased local power variation estimators on a block of adjacent returns. Efficient RNT estimators represent extension of neighborhood truncation estimators theory.

One of the recent works is Mancino and Sanfelici (2012), which proposes new methodology based on Fourier analysis to estimate spot and integrated quarticity. Authors explain that Fourier methodology allows to reconstruct the latent instantaneous volatility as a series expansion with coefficients gathered from the Fourier coefficients of the observable price variation and can be extended to higher even powers of volatility and to the multivariate case. They prove that the Fourier estimator of integrated quarticity is consistent in the absence of noise, then test this new methodology with the use of Monte Carlo experiments and apply it to S&P 500 index futures.

Besides already mentioned papers, which mostly focused on estimation of volatility functionals, there were published other works that provide some useful supplementary methods and theories. For instance, Jacod et al. (2009) presents a generalized pre-averaging approach for estimating the integrated volatility, which also provides consistent estimators of other powers of volatility.

Zhang et al. (2005) analyze in detail different volatility estimators under the presence of market microstructure noise. They also discuss influence of sampling frequency on efficiency of estimators and propose a way of achieving the optimal one under condition of asymptotically small noise.

Method of realized kernels is proposed by Barndorff-Nielsen et al. (2009) for usage with high frequency data to estimate daily volatility of individual stock prices. On addition to that, useful data cleaning procedure are carefully described.

This work aims to provide analysis and test on simulations mentioned above estimators: multipower variation estimator, nearest neighbor truncation estimator and robust neighborhood truncation estimator. To the latter one we pay additional attention and investigate new combinations of this estimator that were not covered by previous papers.

It is divided into four chapters: Chapter 1 contains main theoretical principles of the asset price modeling using continuous-time jump diffusion process, as well as it provides some approaches to jumps detection in the asset price time series and exhibits important role of precise integrated quarticity estimations. Chapter 2 provides theoretical overview of the considered estimators. This part is greatly based on theoretical findings published in Andersen et al. (2009), Andersen et al. (2010) and Andersen et al. (2011), which, when it was possible, were amplified with some additional explanations and calculations. Chapter 3 is devoted to empirical calculations and it starts with description of high frequency data sets used for calculations and cleaning procedures that were applied to eliminate initially error values together with outliers that could possibly distort further estimations. It is followed by illustration of subsampling pre-averaging technique, proposed by Jacod et al. (2009), and concluded by explanations of received empirical integrated quarticity calculations. Finally, Chapter 4 describes conducted simulations and analyses integrated quarticity estimators' efficiency, compares obtained results with those in previous papers, as well as pays additional attention to the usage of robust neighborhood truncation estimators based on group of lower order statistics adjacent log returns. Conclusions provide brief summary of the work and its main results. Appendix contains additional data which covers conducted simulations and empirical results.

1 Generic asset price modeling

1.1 Asset price as a stochastic process

Lets consider normal market trading day which has time length $t \in [0, 1]$. Generally, stochastic process that describes movements of an asset price can be formalized as:

$$dS_t = \mu_t dt + \sigma_t dW_t + dJ_t \tag{1.1}$$

where S_t - is the asset price, μ_t - continuous mean process, σ_t - volatility process, W_t is standard Wiener process and J_t is a finite activity jump process.

Within this time interval $t \in [0, 1]$ we observe n equally spaced logarithmic returns of the asset price $r_i = S_{i/n} - S_{i-1/n}$, $i = 1, \ldots, n$. Under assumption of jump absence $dJ_t = 0$, realized volatility (`RV`) is a consistent estimator for integrated volatility (IV):

$$RV_n = \sum_{i=1}^{n} (S_{i/n} - S_{i-1/n})^2 = \sum_{i=1}^{n} r_i^2 \xrightarrow{\mathrm{P}} IV = \int_0^t \sigma_s^2 ds,\ n \to \infty \tag{1.2}$$

Then for $\mathtt{RV}_n$ holds limiting distribution:

$$\sqrt{n}(RV_n - \int_0^t \sigma_s^2 ds) \xrightarrow{\mathscr{L}} \mathrm{N}(0, 2\int_0^t \sigma_s^4 ds). \tag{1.3}$$

Measure $\int_0^t \sigma_s^4 ds$ is called integrated quarticity , and under mentioned above jump assumption, can be consistently estimated by the realized quarticity (`RQ`):

$$RQ_n = \frac{n}{3} \sum_{i=1}^{n} r_i^4 \xrightarrow{\mathrm{P}} IQ = \int_0^t \sigma_s^4 ds. \tag{1.4}$$

However, under the presence of jumps in the asset price process ($dJ_t \neq 0$), `RV` estimation is no longer consistent for QV, and additionally `RQ` grows indefinitely as far as our sampling frequency becomes greater: $RQ_n \to \infty$, $n \to \infty$. This fact force to estimate integrated volatility and quarticity in other way.

Alternative measure of IV under such circumstances is bipower variation (`BV`):

$$BV_n = \frac{\pi}{2} \left(\frac{n}{n-1} \right) \sum_{i=1}^{n-1} |r_i||r_{i+1}| \tag{1.5}$$

Estimator $\mathtt{BV}_n$ already provides some jump robustness in estimating QV, however, under absence of jumps it is less efficient then RV (Barndorff-Nielsen and Shephard (2006)):

$$\sqrt{n}\left[\int_0^t \sigma^4(s)ds\right]^{-1/2}\begin{pmatrix} RV - \int_0^t \sigma^2(s)ds \\ BV - \int_0^t \sigma^2(s)ds \end{pmatrix} \xrightarrow{\mathscr{L}} \mathrm{N}\left(0, \begin{bmatrix} 2 & 2 \\ 2 & (\frac{\pi}{2})^2 + \pi - 3 \end{bmatrix}\right) \quad n \to \infty. \tag{1.6}$$

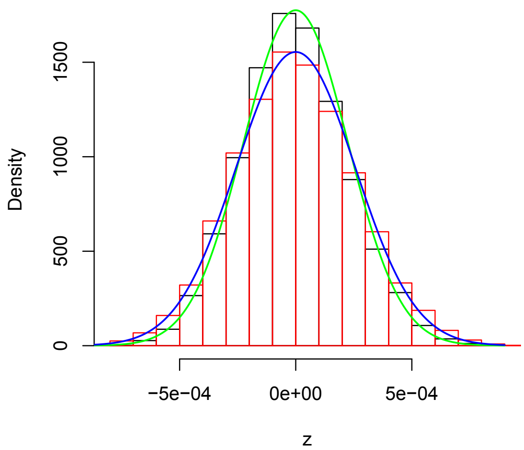

Figure 1.1: Limiting distributions of RV_n and BV_n (black - $z = \sqrt{n}(RV_n - IV)$, red - $z = \sqrt{n}(BV_n - IV)$, green - $\mathrm{N}(0, 2IQ)$) and blue - $\mathrm{N}(0, \left[(\frac{\pi}{2})^2 + \pi - 3\right] IQ)$

1.2 Integrated quarticity as an essential part of jump tests

Appearance of jumps within the price process formulated in Section 1.1 is modeled by component J_t.

Jump process J_t is not continuous and has discrete movements, that are called jumps. While speaking about stock market, asset price jumps represent market's reaction on different news and events happening in the world.

Jumps can be instantaneous (mathematical jump) when stock price S_t in a time moment t changes immediately to the value S_t^+; or it can be gradual, when the stock price is reaching its new level S_t^+ within some time period τ (Barndorff-Nielsen et al. (2009)).

A variety of methods has been develop for identifying jumps along asset price time series and for their separation.

Barndorff-Nielsen and Shephard (2006), for instance, provide us useful expression:

$$\sqrt{n}\frac{(BV - \sum_{i=1}^{n} r_i^2)}{\sqrt{\int_0^t \sigma_s^4 ds}} \xrightarrow{\mathscr{L}} \mathrm{N}(0,\nu), \quad n \to \infty \tag{1.7}$$

which gives us asymptotic distribution of a linear jump statistic $RV - BV$, or similar to this, for a ratio jump statistic $\frac{RV}{BV}$ we have:

$$\sqrt{n}\frac{(BV/\sum_{i=1}^{n} r_i^2 - 1)}{\sqrt{\frac{\int_0^t \sigma_s^4 ds}{\left(\int_0^t \sigma_s^2 ds\right)^2}}} \xrightarrow{\mathscr{L}} \mathrm{N}(0,\nu), \quad n \to \infty \tag{1.8}$$

where $\nu = (\pi^2/4) + \pi - 5$.

These asymptotic distributions give valuable inference about time series variance, however due to the dependence upon the unknown integrated quarticity value $\int_0^t \sigma_s^4 ds$ they are rather statistically infeasible.

As a possible solution, Andersen et al. (2006) use expression:

$$z_{TQ,t} = \sqrt{n}\frac{(BV_t - RV_t)}{\sqrt{(\frac{\pi^2}{4} + \pi - 5)TP_t}} \xrightarrow{\mathscr{L}} \mathrm{N}(0,1), \quad n \to \infty \tag{1.9}$$

for testing the presence of daily jumps. Or alternatively another measure can be used:

$$z_{QQ,t} = \sqrt{n}\frac{(BV_t - RV_t)}{\sqrt{(\frac{\pi^2}{4} + \pi - 5)QQ_t}} \xrightarrow{\mathscr{L}} \mathrm{N}(0,1), \quad n \to \infty, \tag{1.10}$$

which instead of tri-power quarticity uses quad-power QQ.

Based on these test statistics, a couple of other improvements were implemented. In order to boost finite sample performance Andersen et al. (2006) applied logarithms to the variation measures:

$$z_{TQ,t}^{l} = \sqrt{n}\frac{log(BV_t) - log(RV_t)}{\sqrt{(\frac{\pi^2}{4} + \pi - 5)\frac{TQ_t}{BV_t^2}}} \xrightarrow{\mathscr{L}} \mathrm{N}(0,1), \quad n \to \infty, \tag{1.11}$$

$$z_{QQ,t}^{l} = \sqrt{n}\frac{log(BV_t) - log(RV_t)}{\sqrt{(\frac{\pi^2}{4} + \pi - 5)\frac{QQ_t}{BV_t^2}}} \xrightarrow{\mathscr{L}} \mathrm{N}(0,1), \quad n \to \infty. \tag{1.12}$$

All these basic test statistics are quite sensitive to jumps which lets us detect them with high enough preciseness. At the Figure 1.2 you can see an example of *z-test* statistics behavior given the jump presence:

- day without jump (pure BM process)

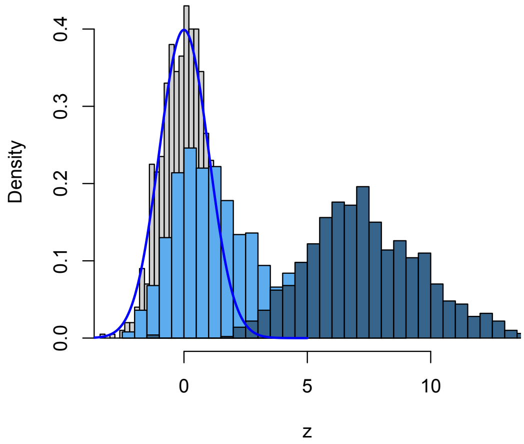

Figure 1.2: Test statistics 1.7 under the influence of jumps with different sizes (blue line - $\mathrm{N}(0,1)$)

- day with a random $0\% - 0.5\%$ jump of the price level;
- day with a random $0.5\% - 0.9\%$ jump of the price level.

Quite significant shift occurs even with such a relatively small presence of discontinuity in the price process. Each of these test statistics depends on IQ measure, thus, imprecise estimation of the latter one will inevitably lead to distorted test results and rising of misclassification rate.

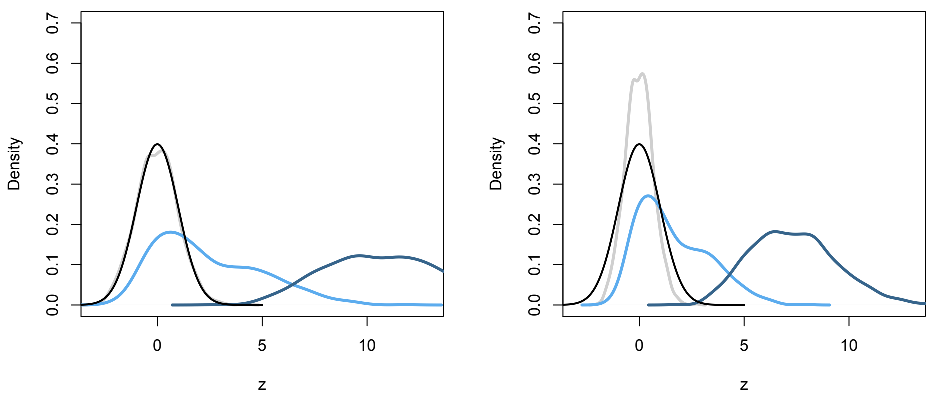

Figure 1.3: Influence of biased quarticity estimate on the jump hypothesis tests

Figure 1.3 illustrates how z-scores can be possibly distorted by wrong estimation of IQ.

At the left-side of the plot, z-statistic from Equation 1.9 is estimated using relatively precise value `TQ` under the influence of different jumps. Right-side plot depicts same statistics and jump sizes, however in this case integrated quarticity value `TQ` was multiplied by random values within the range $2-2.5$ (corresponds to a considerable estimation bias), which has led to the shift of all the z-score curves to the left. Naturally, derivation of integrated quarticity estimators, that will show enough robustness to jumps and microstructure noise, due to more accurate results will positively influence jump tests reliability.

2 Theoretic approaches to quarticity estimates

2.1 Multipower Variation Estimators

Let us have a set of n equally spaced log-returns of the asset price, i.e. $r_i = S_i - S_{i-1},\ i = 1, \dots, n$.

The first class of estimators we are going to highlight, which was proposed by Barndorff-Nielsen Shephard (2002), is the Realized Multipower Variation (MPV). It is defined through the cumulative sum of $n - m + 1$ products of m adjacent absolute log-returns raised to the (p/m)'th power, which guarantees that cumulative power of their product equals p:

$$MPV_n(m,p) \;=\; d_{m,p}\frac{n^{\frac{p}{2}}}{n-m+1}\sum_{i=1}^{n-m+1}|r_i|^{\frac{p}{m}}\cdots|r_{i+m-1}|^{\frac{p}{m}} \xrightarrow{P} \int_0^1 \sigma_s^p ds \tag{2.1}$$

where $m > p/2,\ d_{m,p} = \mu_{p/m}^{-m}$ and $\mu_p = E|U|^p = 2^{p/2}\dfrac{\Gamma((p+2)/2)}{\Gamma(1/2)}, \quad U \sim \mathrm{N}(0,1)$.

Positive integer m sets the window size of return blocks, and p defines the power of the variation, we would like to receive. Term $\left[\dfrac{n}{n-m+1}\right]$ is a finite sample correction factor.

If i.i.d returns are $r_i, ..., r_{i+m-1} \sim \mathrm{N}(0,\sigma^2)$, then $E[|r_i|^{p/m}\cdots|r_{i+m-1}|^{p/m}] \propto \sigma^2$ and thus, after proper normalization, each term of the `MPV` gives an unbiased estimate of the power of spot volatility (Barndorff-Nielsen and Shephard (2004) and Barndorff-Nielsen et al. (2006)). As a result, `MPV` becomes an unbiased, consistent estimator of integrated power variation.

Moreover, using such block-wise structure of adjacent returns for estimating spot volatility provides sufficient jump-robustness to this estimator. In case there is a jump (or several) within a given block, it's contribution will be softened by multiplication with powers of other adjacent returns without jump.

Given different parameters m and p we get `MPV` estimators of lower orders for volatility and quarticity values. In Chapter 1 we have already mentioned bipower variation and bipower quarticity (BQ) is built in analogous way:

$$BV_n = MPV_n(2,2) = d_{2,2}\frac{n}{n-1}\sum_{i=1}^{n-1}|r_i||r_{i+1}|, \tag{2.2}$$

$$BQ_n = MPV_n(2,4) = d_{2,4}\frac{n}{n-1}\sum_{i=1}^{n-1}|r_i|^2|r_{i+1}|^2. \tag{2.3}$$

Also, Andersen et al. (2006) have suggested to use realized tri-power variation (TV) and tri-power quarticity (TQ), which are special cases of MPV estimator:

$$TV_n = MPV_n(3,2) = d_{3,2}\frac{n}{n-2}\sum_{i=1}^{n-2}|r_i|^{2/3}|r_{i+1}|^{2/3}|r_{i+2}|^{2/3}, \tag{2.4}$$

$$TQ_n = MPV_n(3,4) = d_{3,4}\frac{n}{n-2}\sum_{i=1}^{n-2}|r_i|^{4/3}|r_{i+1}|^{4/3}|r_{i+2}|^{4/3}. \tag{2.5}$$

Barndorff-Nielsen and Shephard (2004) have described and analyzed realized quad-power volatility and quarticity estimators:

$$QV_n = MPV_n(4,2) = d_{4,2}\frac{n}{n-3}\sum_{i=1}^{n-3}|r_i|^{1/2}|r_{i+1}|^{1/2}|r_{i+2}|^{1/2}|r_{i+3}|^{1/2}, \tag{2.6}$$

$$QQ_n = MPV_n(4,4) = d_{3,4}\frac{n}{n-3}\sum_{i=1}^{n-3}|r_i||r_{i+1}||r_{i+2}||r_{i+3}|. \tag{2.7}$$

Further, during our analysis we are going to consider MPV estimators up to 5th power.

2.2 Nearest Neighbor Truncation Estimators

Willing to enhance existing MPV estimators, Andersen et al. (2009) proposed new estimators, which they eventually called Nearest Neighbor Truncation Estimators (NNT):

$$MinNNT_n = d_{Min,p}\frac{n}{n-1}\sum_{i=1}^{n-1}[\min(|r_i|,|r_{i+1}|)]^p, \tag{2.8}$$

$$MedNNT_n = d_{Med,p}\frac{n}{n-1}\sum_{i=2}^{n-1}[\mathrm{med}(|r_{i-1}|,|r_i|,|r_{i+1}|)]^p. \tag{2.9}$$

These estimators have better theoretical efficiency properties then MPV estimators plus they also demonstrate greater finite-sample robustness to jumps. The latter is a direct cause of their structure: in case of a jump within a given block of returns, MPV estimator will soften it's contribution by multiplication with powers of other adjacent returns without jump, while NNT with *min* and *med* operators will simply eliminate returns with jumps.

Expressions of NNT estimators for calculating IV values of the time series are following:

$$MinRV_n = d_{Min,2}\frac{n}{n-1}\sum_{i=1}^{n-1}[\min(|r_i|,|r_{i+1}|)]^2, \tag{2.10}$$

$$MedRV_n = d_{Med,2}\frac{n}{n-1}\sum_{i=2}^{n-1}[\mathrm{med}(|r_{i-1}|,|r_i|,|r_{i+1}|)]^2. \tag{2.11}$$

MinRV and MedRV are more efficient then BV estimator due to the smaller bias. As shown in Andersen et al. (2009), if Δj_i is a size of a price jump in the interval $[t_{i-1};t_i]$, and there are no other jumps within adjacent intervals, so that $|\Delta S_{i-1}| << |\Delta j_i|$ and $|\Delta S_{i+1}| << |\Delta j_i|$, thus, within this particular interval, distortion to the overall BV estimate and distortion to the MinRV value satisfy:

$$\frac{\pi}{2}|\Delta j_i|(|\Delta Y_{i-1}| + |\Delta Y_{i+1}) >> \frac{\pi}{\pi-2}\left[|\Delta Y_{i-1}|^2 + |\Delta Y_{i+1}|^2\right]. \tag{2.12}$$

BV distortion on the left hand side is of order $1/\sqrt{n}$, while MinRV on the right hand is of order $1/n$. Moreover, the upward bias for multipower variation estimators $\mathtt{MPV}_n(m;2)$, $m \geq 2$ is of order $1/n^{1-\frac{1}{m}}$, which results in a $1/n$ bias given large m.

Using expression 2.12 authors supported the idea, that not the actual size of jumps matters for MedRV and MinRV but rather their quantity within the interval.

Joint asymptotic distribution between the MedRV, MinRV and estimators RV and BV under the absence of jumps is:

$$\sqrt{n}\begin{bmatrix} RV_n - IV \\ BV_n - IV \\ MinRV_n - IV \\ MedRV_n - IV \end{bmatrix} \xrightarrow{\mathscr{L}} \mathrm{N}\left(0, \begin{bmatrix} 2 & 2 & 2 & 2 \\ & 2.61 & 2.98 & 2.53 \\ & & 3.81 & 3.09 \\ & & & 2.96 \end{bmatrix} IQ\right). \tag{2.13}$$

As can be seen, MinRV estimator under no jump null hypothesis is the least efficient one, while his asymptotic variance is much higher then the others, and MedRV estimator is of comparable efficiency with BV.

Extending further the theory of NNT estimators it is possible to construct estimators, which will cover higher powers of volatility, in particular integrated quarticity:

$$MinRQ_n = d_{Min,4}\frac{n}{n-1}\sum_{i=1}^{n-1}[\min(|r_i|,|r_{i+1}|)]^4, \tag{2.14}$$

$$MedRQ_n = d_{Med,4}\frac{n}{n-1}\sum_{i=1}^{n-1}[\mathrm{med}(|r_{i-1}|,|r_i|,|r_{i+1}|)]^4. \tag{2.15}$$

Asymptotic theory for these estimators is quite similar to the realized volatility estimators. Thus according to Andersen et al. (2010):

$$\sqrt{n}(MinRQ_n - IQ) \xrightarrow{\mathscr{L}} \mathrm{N}\left(0, 18.54 \int_0^1 \sigma_s^8 ds\right) \tag{2.16}$$

$$\sqrt{n}(MedRQ_n - IQ) \xrightarrow{\mathscr{L}} \mathrm{N}\left(0, 14.16 \int_0^1 \sigma_s^8 ds\right) \tag{2.17}$$

$$\tag{2.18}$$

Here `NNT` estimator that uses *min* function is still less efficient then `MinRQ`, demonstrating higher level of asymptotic volatility.

Willing to see a wider picture that will include `TQ`, `MPQ4`, `MPQ5` estimators, we have calculated asymptotic joint distribution for seven different estimators under no jump hypothesis:

$$\sqrt{n}\begin{bmatrix} RQ_n - IQ \\ BQ_n - IQ \\ MinRQ_n - IQ \\ MedRQ_n - IQ \\ TQ_n - IQ \\ MPQ4_n - IQ \\ MPQ5_n - IQ \end{bmatrix} \xrightarrow{\mathscr{L}} \mathrm{N}\left(0, \begin{bmatrix} 10.73 & 8.00 & 6.96 & 7.27 & 7.26 & 6.81 & 6.55 \\ & 12.06 & 13.57 & 11.44 & 11.71 & 11.13 & 10.69 \\ & & 18.19 & 14.61 & 13.09 & 12.29 & 11.71 \\ & & & 14.06 & 11.15 & 10.62 & 10.23 \\ & & & & 13.91 & 13.89 & 13.61 \\ & & & & & 15.12 & 15.27 \\ & & & & & & 16.11 \end{bmatrix} \int_0^1 \sigma_s^8 ds\right). \tag{2.19}$$

These values are quite approximate, but they definitely show the proper scale and can give the notion about relative efficiency of each of the estimators against others. Thus, considering asymptotic volatility values, `MinRQ` estimator stands out like the least efficient one and most surely reaches the asymptotic volatility levels of `MPQ6` and `MPQ7`. `MedRQ` outperforms `MinRQ` and estimators with powers higher then `TQ`, however still losing to `RQ` and `BQ`.

Nearest neighbor truncation estimators `MinRQ` and `MedRQ` operate with compact return blocks which gives them enough robustness towards time variation in volatility, plus their functional structure provides enough jump-resistance. These features make them competitive with the other estimators and makes possible their practical implementation.

2.3 Robust Neighborhood Truncation Estimators

Nearest neighbor truncation estimators `MinRQ` or `MedRQ` are part of the far more general class - Neighborhood Truncation Estimators (`NT`). Proposed for the first time in the paper Andersen et al. (2011), they represent efficient jump-robust approach to estimating integrated quarticity.

Let $r_i = S_{i/n} - S_{i-1/n}$, $i = 1, \ldots, n$ be n equally spaced logarithmic returns of the asset price. Then, we denote ith block of absolute returns as $\underline{r}_{i,m} = (|r_i|, ..., |r_{i+m-1}|)$, $i =$

$1, \ldots, n-m+1$ and jth order statistic of the ith absolute return block as $q_j(|r_1|, \ldots, |r_m|) = q_j(\underline{r}_{i,m})$, naturally $q_1(\underline{r}_{i,m}) \leq \ldots \leq q_m(\underline{r}_{i,m})$.

Following these notations, baseline Neighborhood Truncation estimator ($\mathtt{NT}_n^{(j,m)}(p)$) is given by

$$NT_n^{(j,m)}(p) = d_{(j,m)}(p)\left(\frac{n^{p/2}}{n-m+1}\right)\sum_{i=1}^{n-m+1}\left[q_j(\underline{r}_{i,m})\right]^p, \quad j = 1, \ldots, m, \tag{2.20}$$

where $d_{(j,m)}(p) = \{\mathsf{E}[q_j(|Z_1|^p, \ldots, |Z_m|^p)]\}^{-1}, \quad Z_i \sim \ i.i.d.\ \mathrm{N}(0,1),\ i = 1, ...m.$

Basically, NT estimator is a properly scaled sum of jth order statistics of ith absolute return block, raised to pth power. Placing scaling factor $d_{(j,m)}(p)$ in front of pth power of the jth absolute order statistic gives an unbiased estimator for σ^p. As can be seen, $\mathtt{MinPV}(p)$ is $\mathtt{NT}_n^{(1,2)}(p)$ with a scaling factor $d_{(1,2)}(p)$ and $\mathtt{MedPV}(p)$ is $\mathtt{NT}_n^{(2,3)}(p)$ with $d_{(2,3)}(p)$.

Robust neighborhood truncation estimator (RNT) represents further extension of this approach, which results in higher jump-robustness and efficiency.

General algorithm, proposed by Andersen et al. (2011), is defined as:

$$RNT_n^{(j,\boldsymbol{I})}(p) \ = \ d_{(j,\boldsymbol{I})}(p)\frac{1}{n-m+1}\sum_{i=1}^{n-m+1} q_j\left[\varepsilon_{k_1}(\underline{r}_{i,m}), \ldots, \varepsilon_{k_H}(\underline{r}_{i,m})\right] \tag{2.21}$$

where

$$\varepsilon_{k_H} \ = \ d_{(k_H,m)}(p)n^{p/2}\left[q_{k_H}(\underline{r}_{i,m})\right]^p. \tag{2.22}$$

Firstly, within the given ith return block we calculate properly scaled functional of needed order statistics $\varepsilon_{k_1}(\underline{r}_{i,m}), \ldots, \varepsilon_{k_H}(\underline{r}_{i,m})$. Vector $\boldsymbol{I} = \ (k_1, \ldots, k_H), \quad 1 \leq H \leq m$ in this case defines which combination of order statistics we would like in each concrete return block. After that, to the received set of H unbiased estimators of σ^p $\{\varepsilon_{k_1}, \ldots, \varepsilon_{k_H}\}$ we apply jth order statistics, which is scaled by respective factor $d_{(j,\boldsymbol{I})}(p)$. This gives us final value of return functional for the ith return block.

RNT estimator is consistent and propositions stated by Andersen et al. (2011) are valid:

$$RNT_n^{(j,\boldsymbol{I})}(p) \xrightarrow{\mathrm{P}} \int_0^1 \sigma_s^p ds, \quad j = 1, \ldots, H \tag{2.23}$$

and, given volatility process follows generalized Itô process

$$\sqrt{n}\left(RNT_n^{(j,\boldsymbol{I})}(p) - \int_0^1 \sigma_s^p ds\right) \xrightarrow{\mathscr{L}} \mathrm{N}\left(0, \eta(j,\boldsymbol{I};p)\int_0^1 \sigma_s^{2p} ds\right), \quad j = 1, \ldots, H, \tag{2.24}$$

where $\eta(j,\boldsymbol{I};p)$ is a known constant.

Naturally, the $d_{(j,\boldsymbol{I})}(p)$ scaling factor, which converts jth order statistics, applied to the set of unbiased σ^p estimators $\varepsilon_{k_1}(\underline{r}_{i,m}), \ldots, \varepsilon_{k_H}(\underline{r}_{i,m})$, into a robust unbiased estimator of the given ith return block, depends on the initial configuration of the set:

$$d_{(j,\boldsymbol{I})}(p) = \left\{\mathsf{E}\left[q_j(d_{(k_1,m)}(p)Z^p_{(k_1,m)}, \ldots, d_{(k_H,m)}(p)Z^p_{(k_H,m)})\right]\right\}^{-1}, \quad (2.25)$$

$$d_{(k_h,m)}(p) = \{\mathsf{E}[q_{k_h}(|Z_1|^p, \ldots, |Z_m|^p)]\}^{-1}, \quad Z_i \sim \ i.i.d.\ \mathrm{N}(0,1), \quad (2.26)$$

$$\boldsymbol{I} = (k_1, \ldots, k_H), \quad 1 \leq H \leq m. \quad (2.27)$$

Since, usually there are no closed form solutions for the $d_{(j,\boldsymbol{I})}(p)$ coefficients, their values are obtained via simulations. Let us consider `RNT` estimators with return block size $m = 6$

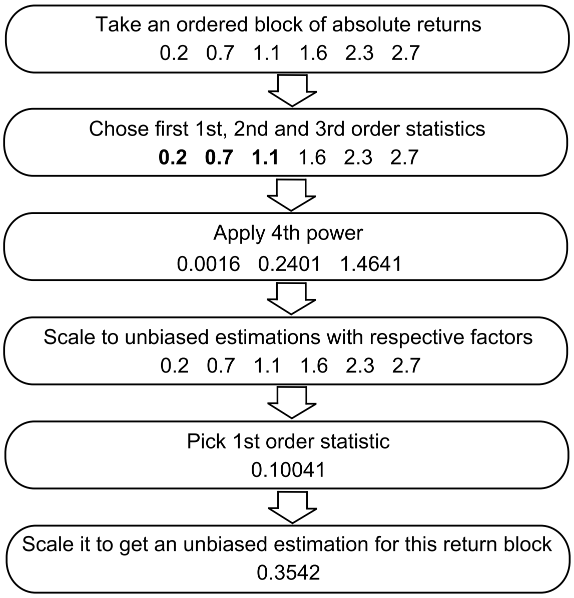

Figure 2.1: RNT6 1(123) estimator of σ^4 based on lower order statistics returns

and power $p = 4$, namely `RNTQ6 1(123)`, `RNTQ6 2(123)`, `RNTQ6 1(456)`, `RNTQ6 2(456)` (letter Q is added due to the fact that power $p = 4$ produces quarticity values). For convenience further on, estimators based on returns `(123)` or, say, `(1234)` we will call lower order statistics `RNTQ` estimators (shortly `LOS RNTQ`) and `(456)` or `(4567) -` higher order statistics `RNTQ` estimators (`HOS RNTQ`). During first simulations we will compare efficiency of these estimators with other non `RNT` measures under the influence of random jumps and other market data imperfections. Scaling factors we have calculated for each

	$d_{(1,3)}(4)$	$d_{(2,3)}(4)$	$d_{(3,3)}(4)$	$d_{(j,\boldsymbol{I})}(4)$
RNTQ6 1(123)	62,75698	10,88057	2,96839	3,52776
RNTQ6 2(123)	63,24496	10,81908	2,94555	1,29788
RNTQ6 1(456)	0,95240	0,30849	0,07552	2,32949
RNTQ6 2(456)	0,95164	0,30800	0,07540	1,17506

Table 2.1: Scaling factors of RNTQ6 estimator for different order statistics

of these estimators separately and, in order to achieve needed preciseness, within each of the simulations a 1 mln. repetitions were performed. Values that were received and later used to calculate all `RNT6Q` estimators are illustrated in Table 2.1. They vary quite significantly: for statistics of lower orders they are biggest and then they diminish as far as we take statistics of higher orders.

Similar to illustration provided in Andersen et al. (2011) for their `RNTQ5 1(345)` estimator, with a Figure 2.1 we would like to clarify the structure of `RNTQ6 1(123)` which uses lower order statistics.

3 High frequency data preparation

3.1 Data aggregation and filtering procedures

Empirical part of our work covers application of quarticity estimators to real market data. Estimators that we have already discussed seem to rely heavily on such time series features as frequency and data regularity. While these parameters are not necessarily correlated with trading volumes, still latter can somehow help us to differentiate companies, thats why we were choosing stocks by volume amounts: *Susquehanna Bancshares Inc.* (Ticker: *SUSQ*), *Pfizer Inc.* (*PFE*), *Exxon Mobil Corporation* (*XOM*).

Limit order book data based on NASDAQ's historical ITCH database was provided by `LOBSTER` system[1] and covered timespan from 1st of July 2008 till 17th of November 2010. From some reason data set did not contain data for end of November and full December month.

Each of the trading days was described by 2 data files: Message File which contained trade data and Orderbook File with quote data. Small excerpts of these datasets are provided in Tables 3.1 and 3.2, respectively.

Time (ms)	Ask Price 1	Ask Size 1	Bid Price 1	Bid Size 1
32409050	2463300	50	2461200	200
32414175	2463300	50	2461700	100
32428842	2462100	8	2461700	100
32430926	2463300	50	2461700	100
32441810	2463300	50	2461700	10
32443298	2462500	100	2461700	10
32444021	2463300	50	2461700	10
32444598	2462500	100	2461700	10
32444912	2462400	100	2461700	10
32445642	2463300	50	2461700	10

Table 3.1: Abstract of the Orderbook data file

Chosen time period contained almost 600 days, thus totally it was covered by roughly 1200 data files with quotes and trades. After merging both of these tables for each of the

[1] The project developed at the Chair of Econometrics at the Humboldt-Universität zu Berlin in cooperation with Research Data Center of the Collaborative Research Center 649: Economic Risk. More information is available by the link `http://lobster.wiwi.hu-berlin.de/`

Time (ms)	Type	Message ID	Size	Price	Direction
32409050	1	3442887	100	2461200	1
32414175	1	3446221	100	2461700	1
32428842	1	3458996	8	2462100	-1
32430926	4	3458996	8	2462100	-1
32441810	4	3446221	90	2461700	1
32443298	1	3468540	100	2462500	-1
32444021	3	3468540	100	2462500	-1
32444598	1	3469276	100	2462500	-1
32444912	1	3469524	100	2462400	-1
32445642	3	3469524	100	2462400	-1

Table 3.2: Abstract of the Message data file

days, precise cleaning and data filtering was an important step before volatility estimations.

Firstly, since all the timestamps were measured in milliseconds after midnight we converted them to more appropriate 9:30-16:00 hour time format.

Initially, all transactions present in our datasets were classified by types:

- 1: Submitted new order
- 2: Cancellation (Partial deletion of an order)
- 3: Deletion (Total deletion of an order)
- 4: Execution (Against visible order)
- 5: Execution (Against hidden order)
- 100: Other (Unknown)

Among them we have picked only executed transactions (Type 4 and 5), while others were rather technical indicators and were not involved in the price process formation. Columns, such as `Message ID` and `Direction` were deleted due to no need. All traded prices and bid-ask prices were scaled to meet the Euro.Cent format.

Additionally we have considered filtering approaches proposed in the paper Barndorff-Nielsen et al. (2009). After combining our data corrections and mentioned filtering methods, we received following sequence of steps:

- **All data.**

  [A1] delete entries with timestamp outside the 9:30-16:00 window when the exchange is open;

  [A2] omit rows with bid, ask or transaction price equal to zero;

- **Trade data.**

  [T1] choose only transactions with executed orders (visible and hidden with Type 4 and Type 5);

  [T2] within one timestamp, substitute several trades with one using median price;

  [T3] delete entries with prices that are above the ask plus the bid-ask spread and prices below the bid minus the bid-ask spread;

- **Quote data.**

  [Q1] when the multiple quotes have the same time execution, replace them with a single entry with median bid and ask prices;

  [Q2] delete rows for which the spread is negative;

  [Q3] delete entries for which the spread is more than 10 times the median spread on that day;

  [Q4] delete rows for which the mid-quote deviated by more than 10 mean absolute deviations from a centered median (excluding the observation under consideration) of 50 observations;

Table 3.3 summarizes quantities of omitted data at each of the cleaning steps plus final data samples.

Stock	Year	A1	A2	T1	T2, Q1	T3	Q2	Q3	Q4	Clean
SUSQ	2008	0	0	9395270	352196	0	0	530	9	253935
	2009	1	0	11443588	600255	0	0	1004	40	316195
	2010	0	0	7300849	437412	0	0	344	5	198624
PFE	2008	0	0	53919872	5465855	0	0	134	20	683588
	2009	173191	0	63963281	7082214	0	0	51	210	768356
	2010	65389	0	45195467	3888713	0	0	2	157	419683
XOM	2008	0	0	86130414	10744238	0	0	1494	0	1627514
	2009	280757	0	126867580	9501768	0	0	1390	4	1845293
	2010	98451	0	68390341	5645230	0	0	754	1	1074818

Table 3.3: Amounts of omitted data during each of the cleaning steps and final quantity

Easy to notice that rules $A2$, $T3$ and $Q2$ did not influence our sample: all the columns contain zeros. This gives us a hint that initial data did not contain any crucial errors such as zero prices or wrong bid-ask quotes.

While *SUSQ* stock data was not influenced that much by rule $A1$, stocks *PFE* and *XOM* did lose some data points, which have fallen out of 9:30-16:00 time window.

Step $T1$ in our case eliminates most of the data entries due to deleting all prices and quotations except those with Types 4 and 5.

Second biggest amount of deleted values was after applying rules $T2$ and $Q1$. Usually, stocks with higher liquidity are traded more often, which causes rise in transaction number during some period of time and inevitably leads to rising amount of transactions that receive same timestamps. This fact is also reflected in our results: amount of deleted values grows from less active SUSQ stock to more actively traded XOM.

Finally, rules $Q3$ and $Q4$ dropped out relatively significant quantity of outliers (mostly *SUSQ* and *XOM* were influenced).

	Price	Size	Trades	Bid	BidSize	Ask	AskSize
04.01.2010 9:30:56	68.740	1100	11	68.730	3800	68.750	6320
04.01.2010 9:30:57	68.750	400	4	68.740	1000	68.760	400
04.01.2010 9:30:59	68.760	1300	10	68.760	2400	68.770	9500
04.01.2010 9:31:02	68.760	1100	11	68.760	3145	68.770	2492
04.01.2010 9:31:04	68.725	700	6	68.725	1400	68.740	4092
04.01.2010 9:31:05	68.710	200	2	68.690	600	68.715	1300
04.01.2010 9:31:06	68.720	250	3	68.710	956	68.730	850
04.01.2010 9:31:08	68.750	2200	16	68.740	2670	68.760	7700
04.01.2010 9:31:09	68.740	503	3	68.740	728	68.750	997
04.01.2010 9:31:10	68.740	300	3	68.730	500	68.750	300

Table 3.4: Abstract of the aggregated clean data set

After running these filtering procedures for each of the three stocks we have received clean data, whose format is more convenient for further computations (Table 3.4). Roughly speaking, average daily number of transactions for *Susquehanna Bancshares* was around 1280, for *Pfizer* - 3120 and for *Exxon Mobil* - 7580. Thus, taking into account time length of an ordinary trading day equal 23400 seconds, average times of transaction arrivals were 18.3, 7.5 and 3.1 seconds respectively, which provides us useful range of frequencies.

3.2 Eliminating microstructure noise using pre-averaging technique

In a recent econometric literature it is quite widely accepted that the true price process and the true return data are contaminated by market microstructure effects, e.g. price discreteness and bid-ask spreads, which cause observed asset prices diverge from their efficient values (Bandi and Russell (2003)). The more high frequent is the data, the more exposed it is to the microstructure noise, which inevitably leads to biased estimations.

Useful method, that helps to lower this negative influence to some extent, has been examined thoroughly analyzed in the paper Jacod et al. (2009), and shortly reviewed in Andersen et al. (2011).

Let's have n equispaced returns $r_i = S_i - S_{i-1}$, $i = 1, \dots, n$. Pre-averaged returns with a window size $2k \le n$ are defined as:

$$\overline{r}_i = \frac{1}{k}\sum_{j=k}^{2k-1} S_{i+j} - \frac{1}{k}\sum_{j=0}^{k-1} S_{i+j}, \quad i = 1, \dots, n-2k+1. \tag{3.1}$$

This calculation gives a smoother return paths with less noise (Figure 3.1).

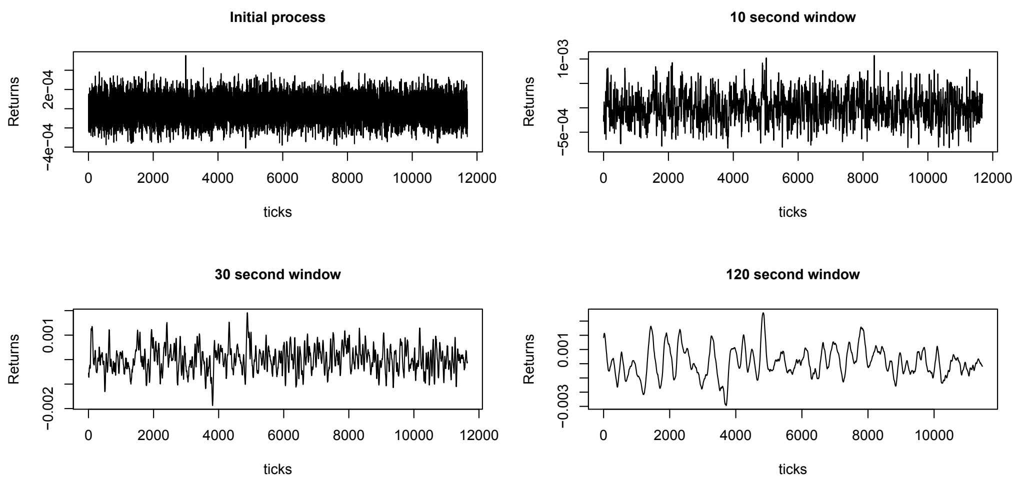

Figure 3.1: Log-returns and pre-averaged returns with different window sizes

Further, out of pre-averaged returns we consider $2k$ subsamples:

$$\begin{aligned}
\overline{P}_1 &= \left\{\overline{r}_{1+2(p-1)k} \ : \ s = 1, \dots, \left\lfloor \frac{n}{2k} \right\rfloor\right\}, \\
\overline{P}_2 &= \left\{\overline{r}_{2+2(p-1)k} \ : \ s = 1, \dots, \left\lfloor \frac{n-1}{2k} \right\rfloor\right\}, \\
&\dots \\
\overline{P}_{2k} &= \left\{\overline{r}_{2k+2(p-1)k} \ : \ s = 1, \dots, \left\lfloor \frac{n-2k+1}{2k} \right\rfloor\right\}.
\end{aligned}$$

We calculate integrated volatility and quarticity estimates for each of the subsamples $\overline{P}_i$ and construct $\widehat{IV}_{\overline{r}}$, $\widehat{IQ}_{\overline{r}}$ estimators for the full set of pre-averaged returns $\overline{r} = \{\overline{r}_i\}_{i=1}^{n-2k+1}$:

$$\begin{aligned}
\widehat{IV}_{\overline{r}} &= \frac{1}{2k}\sum_{i=1}^{2k} \frac{1}{\psi}\widehat{IV}_{\overline{P}_i}, \\
\widehat{IQ}_{\overline{r}} &= \frac{1}{2k}\sum_{i=1}^{2k} \left(\frac{1}{\psi}\right)^2 \widehat{IQ}_{\overline{P}_i},
\end{aligned}$$

where $\psi_k = \dfrac{1}{2k}\displaystyle\sum_{j=1}^{2k-1} 4f(\frac{j}{2k})^2, \ f(x) = x \wedge (1-x), \ x \in [0,1].$ It is a finite sample analog of a variance scaling factor, which, given different window size k, will receive slightly greater values than $\frac{1}{3}$, while in general $\psi = \int_0^1 4f(u)^2 du = \frac{1}{3}$.

Due to finite sample, after applying bias correction we receive the final expressions for our pre-averaged estimators:

$$\begin{aligned}\widehat{IV}_{\overline{r}}^{Adj} &= \frac{1}{2k}\sum_{i=1}^{2k}\frac{1}{\psi}\frac{n/2k}{\lfloor (n-i+1)/(2k)\rfloor}\widehat{IV}_{\overline{P}_i},\\ \widehat{IQ}_{\overline{r}}^{Adj} &= \frac{1}{2k}\sum_{i=1}^{2k}\left(\frac{1}{\psi}\frac{(n/2k)}{\lfloor (n-i+1)/(2k)\rfloor}\right)^2\widehat{IQ}_{\overline{P}_i}.\end{aligned}$$

These estimates demonstrate good noise-robustness and at the same time remain consistent and asymptotically normal. Basically, all of the IQ values we are going to use further for our analysis and comparison will be corrected with this method.

3.3 Empirical calculations based on real market data

Willing to have a glance on different industries and stocks with various trading frequencies we have picked, as was mentioned in Section (3.1), three different stocks: Susquehanna Bancshares Inc. (company that provides a range of retail and commercial banking and financial services), Pfizer Inc. (research-based, global biopharmaceutical company), Exxon Mobil Corporation (manufacturer and marketer of commodity petrochemicals).

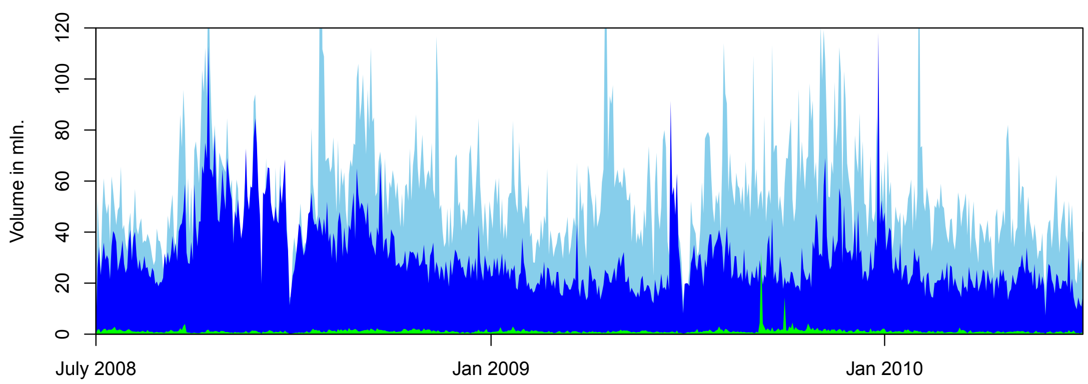

Figure 3.2: Volume dynamics of the companies' stocks: *Pfizer* - cyan, *Exxon Mobil* - blue, *Susquehanna Bancshares* - green

Each of these companies is characterized by different trading volumes. While *Susquehanna Bancshares* has relatively small amounts of daily volumes, *Pfizer* and *Exxon Mobil* are much more heavily traded stocks (Figure 3.2). Worth to mention, trading volumes of each single stock in our case, actually, do not represent the frequency of transactions. Thus, *Pfizer*, having almost twice greater average daily volume then *Exxon Mobil* (around

55 mln. stocks against 31 mln.), in a matter of fact was traded with way less transactions (Table 3.3). This can be explained by the fact that *Exxon Mobil* is more expensive stock then *Pfizer* and it has smaller average size of a single executed transaction: almost 1745 stocks against 3252 at *Pfizer*.

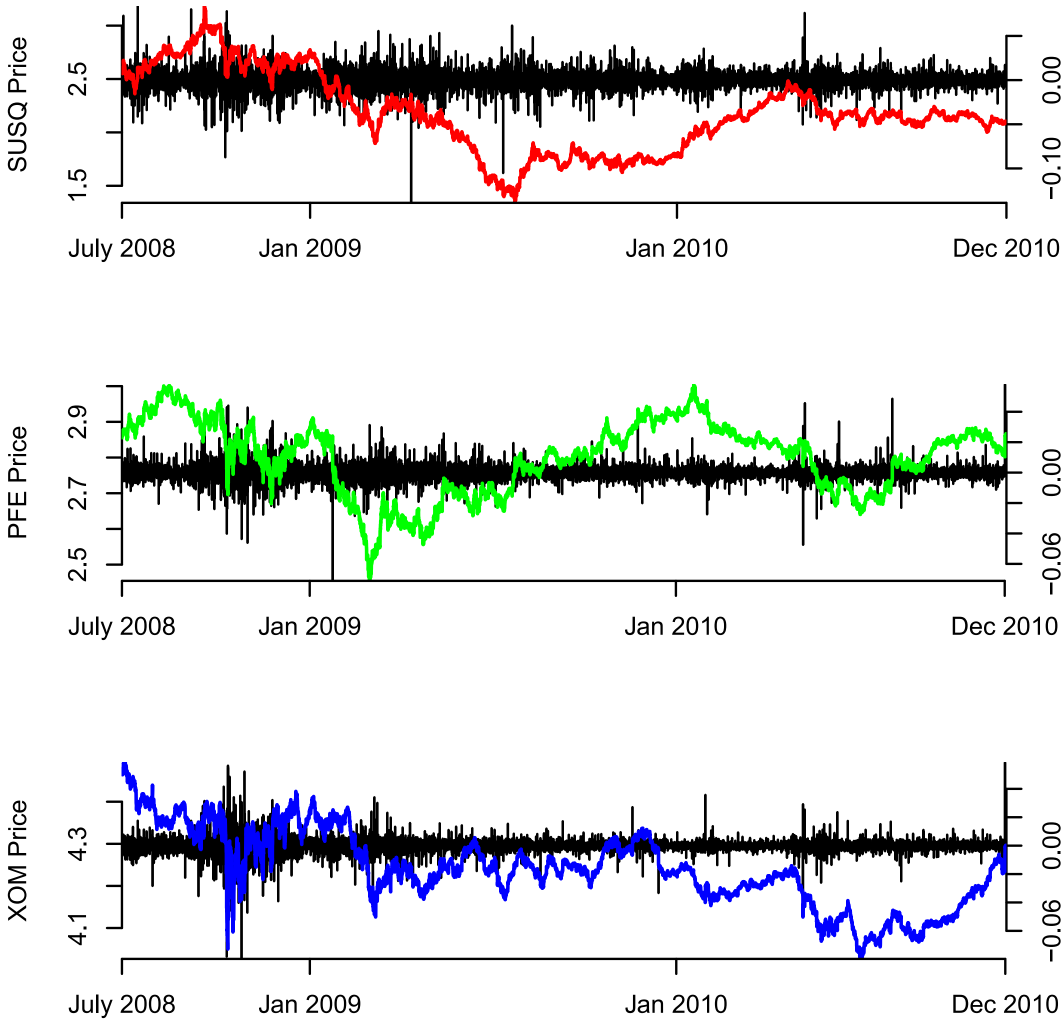

Figure 3.3: Price dynamics of companies' stocks

Time period, we have chosen for estimations, includes second half of the year 2008 and beginning of 2009 - exactly that time span when the recent economical crisis has been shacking world financial markets most furiously. During this time company stocks were demonstrating significant rise in volatilities, increased sizes of price movements and as a result log-returns (Table 3.3).

It can be observed, that log-returns of *SUSQ* stock overall have greatest magnitude among these three, and *XOM* log-returns sizes are relatively lowest ones. One of the possible explanations of this fact is possibly hidden behind the market capitalization of the company and stock's trading frequency. The smaller is the size of the company or the less liquid are its stocks, the greater returns investors demand in order to compensate their inconveniences and exposure to greater market risks. This effect, which in literature is commonly referred to as the "small firm effect", was discussed and analyzed in a variety of papers, e.g. Drew et al. (2006).

Further on, empirical quarticity values for the year 2008 and 2009-2010 we have decided to plot separately, while in the first period they were incomparably greater then in the second one. All estimators were grouped pairwise by their respective classes and results were illustrated with Figure (3.4), Figure (3.5) and Figures (1-4) presented in Appendix. Estimators at the left side of the captions were always plotted with blue color and ones at right-side with black. Aiming to make graphics visually comparable to each other, we have set up following ordinate axis limits: *SUSQ* 2008: $(0, 300e-06)$, *PFE* 2008: $(0, 100e-06)$, *XOM* 2008: $(0, 100e-06)$; *SUSQ* 2009-2010: $(0, 100e-06)$, *PFE* 2009-2010: $(0, 10e-06)$, *XOM* 2009-2010: $(0, 5e-06)$.

As was just mentioned, during September-December 2008 all the stocks are characterized by a limited group of substantial price shocks which resulted in presence of big quarticity "spikes". Their sizes are so significant, that all other data points around seem to have little or no price activity. Differences in estimators are directly driven by magnitude of the price jumps - the greater is the latter one the more similar results the `IQ` estimators give. Within this time period, pairs of estimators `RNTQ6 1(123)`, `RNTQ6 1(456)` and `RNTQ6 2(123)`, `RNTQ6 2(456)` produced a bit lower quarticity values then pairs `MPV(3,4)`, `MPV(4,4)` and `MinRQ`, `MedRQ`, thus showing superior robustness to such big jumps.

From beginning of 2009, quarticity measures started to look a little bit more like a process with some stochasticity. The scale of quarticity measures for each stock naturally mirrors the magnitudes of their log-returns. Given the scale we have chosen, *SUSQ* has the most volatile structure, rich of instantaneous sharp peaks, while stocks *PFE* and *XOM* have considerably lower values along all the period.

Estimators `RNTQ` continue to give lower values then all the other estimators and most of the jumps are significantly dampened. In particular, `LOS RNTQ` such as `RNTQ6 1(123)` and `RNTQ6 2(123)`, appeared to be the most robust to jump presence.

In both of the pairs `MPV(3,4)`, `MPV(4,4)` and `MinRQ`, `MedRQ`, latter estimators `MPV(4,4)` and `MedRQ` were more jump robust then first ones, and in such a way supported the theory worded in Chapter 2.

While proposed empirical part gives us some valuable notion of estimators' behavior during application to real life data sets, we cannot distinguish on its basis estimators with superior efficiency. In such case, conduction of Monte Carlo simulations will let us scrutinize each of the mentioned measures under the influence of various market imperfections.

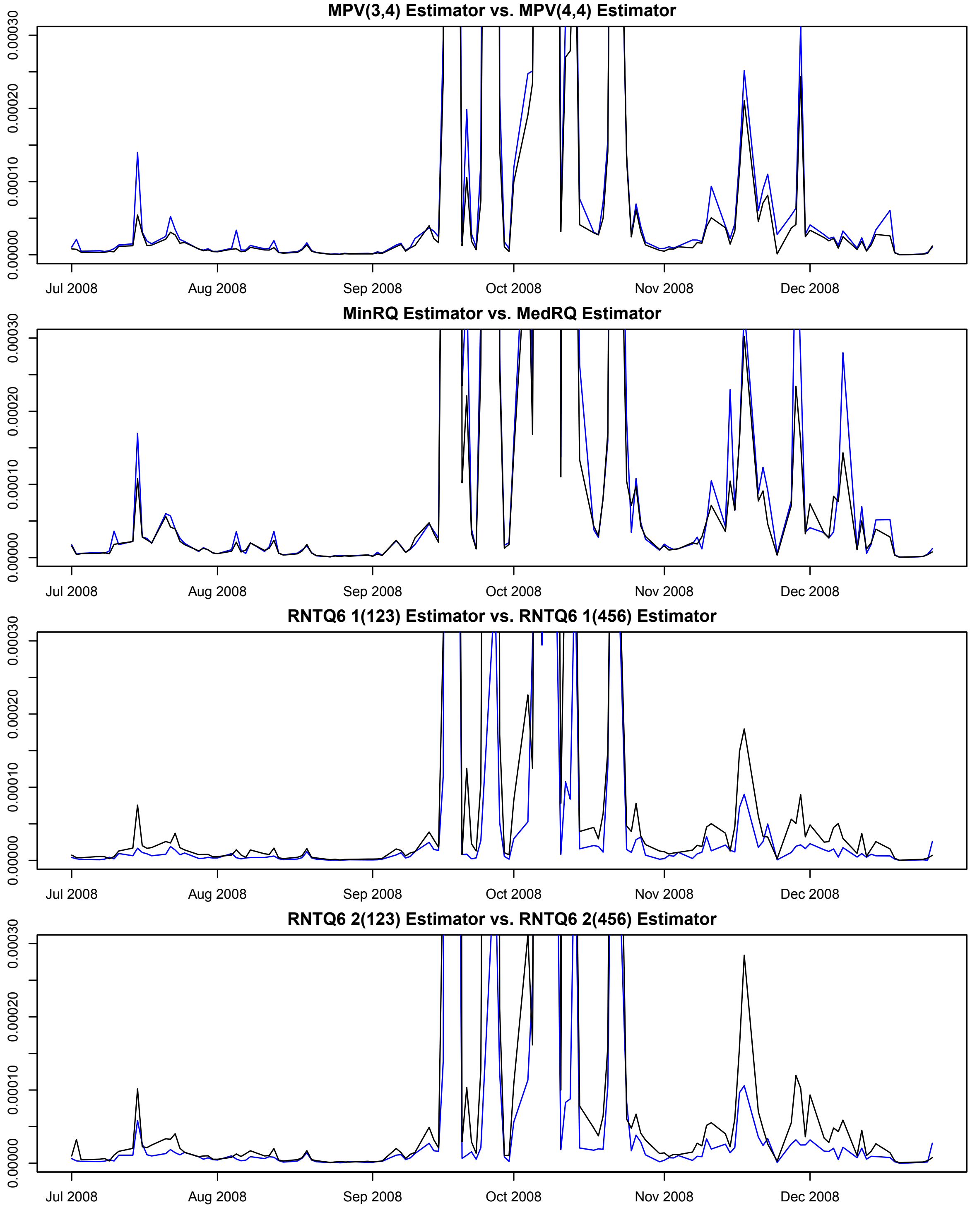

Figure 3.4: Integrated quarticity estimations for Susquehanna Bancshares during year 2008

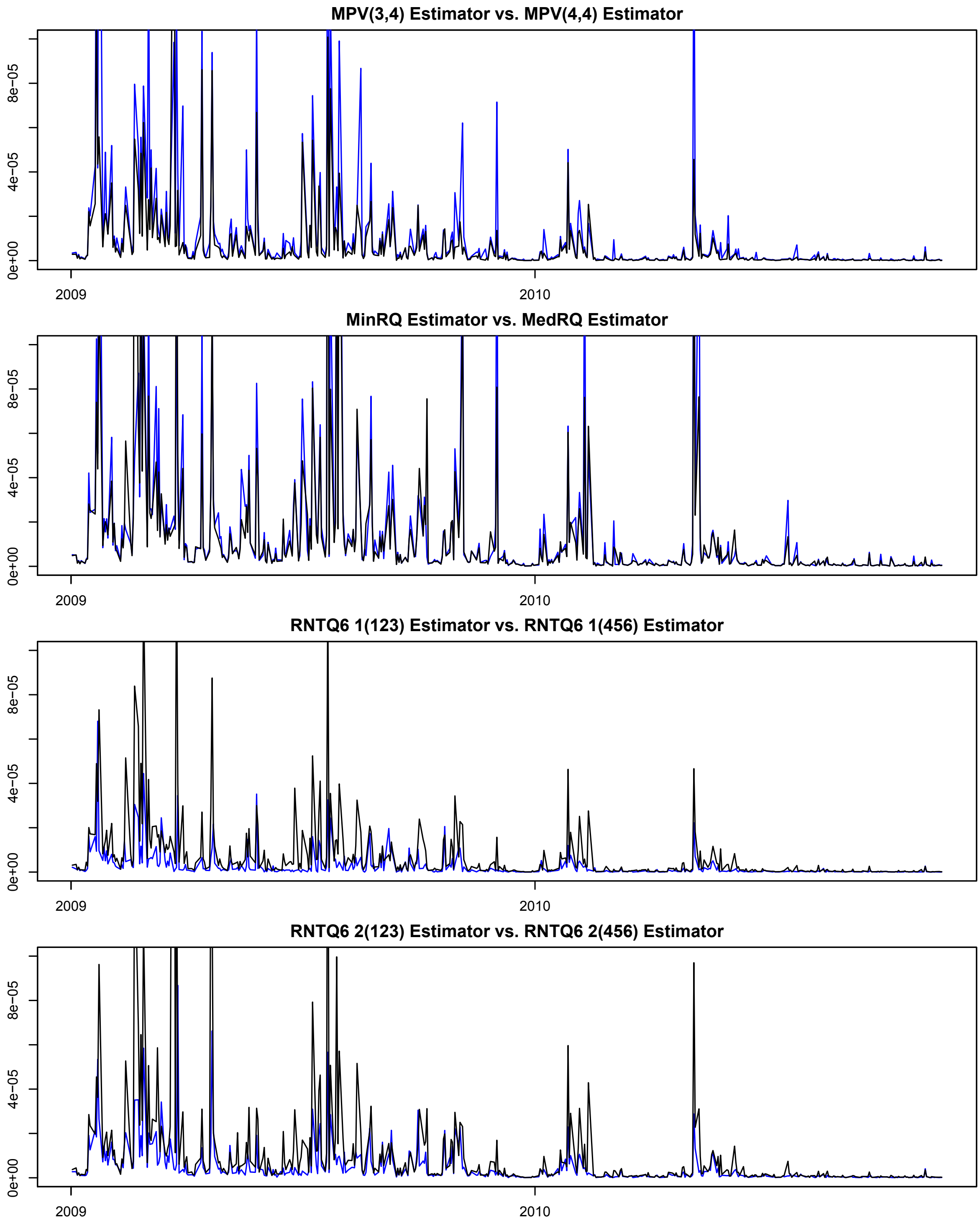

Figure 3.5: Integrated quarticity estimations for Susquehanna Bancshares during years 2009-2010

4 Benchmarking quarticity estimators via simulations

In this chapter we focus on variate Monte Carlo simulations in order to see how different patterns of stochastic asset price process can influence performance of our quarticity estimators. Given each of the models, we will subsample with different time intervals in order to capture the influence of different sampling frequencies.

At the beginning we decided to examine such estimators: `RQ`, `TQ`, `MPQ4`, `MPQ5`, `MinRQ`, `MedRQ`, `RNTQ6 1(123)`, `RNTQ6 2(123)`, `RNTQ6 1(456)` and `RNTQ6 2(456)`.

Estimation models, we are going to use, were also examined in Andersen et al. (2009), Andersen et al. (2010), as well as in Andersen et al. (2011) and Podolskij and Vetter (2006):

- Brownian motion simulation (BM) with and without jumps;
- Stochastic volatility model with intraday U-shape volatility pattern (SV-U model)
- Sparse sampling model (irregular trade intervals).

Within all the models (except sparse sampling), we simulate data between 9:30 and 16:00 with a 1 second interval, which results in 23400 observations per day. For the sparse sampling model another approach is used, and will be mentioned later in Section 4.4. The unconditional daily volatility is set to 0.000159, which is equivalent to around 20% per annum.

In each of the cases 2400 days were simulated, which covers almost 10 years of stock market activity. All simulations were coded in statistical programming language `R` and, due to big amounts of data calculations, we have applied the `SNOW` package[1] which provides elegant solutions for multi-processor parallel computing. This let us utilize 6 cores, each of them performing 400 independent iterations.

[1]Additional information and resources are provided at `http://cran.r-project.org/web/packages/snow/index.html`

4.1 Brownian motion process with and without jumps

Our first model will be standard Brownian motion, which represents ideal scenario of possible price movement:

$$dS_t = S_t \sigma dW_t.$$

Given this framework we receive continuous stochastic process without any price jumps, which, supported by the asymptotic theory stated in Chapter 1, should cause all estimators of IV and IQ be unbiased and consistent.

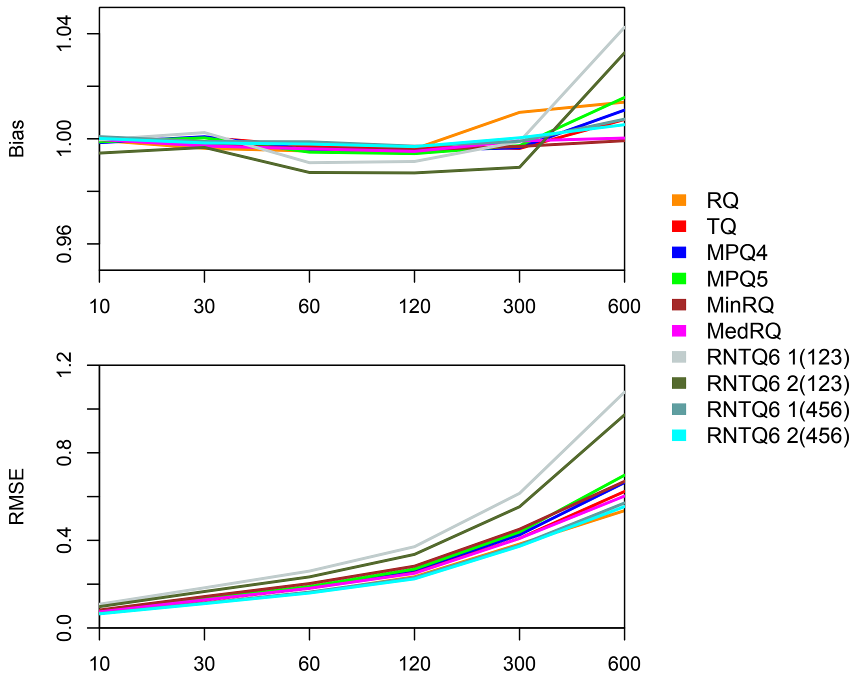

Figure 4.1: Brownian motion asset price simulation

Indeed, at the Figure 4.1 we see that almost all of the estimators have a significantly small or no bias. Proposed estimators `RNTQ6 1(123)` and `RNTQ6 2(123)` have slightly greater downward bias in range of 30-300 seconds and after 300 second window size.

All estimators have almost identical Realized Mean Square Errors (RMSE) that rise gradually with a rising sampling frequency. Estimators `RNTQ6 1(123)` and `RNTQ6 2(123)` have higher values then the others.

As was mentioned before, one of the crucial properties of IQ estimators, is their robustness to jumps. Thus, next logical step is adding some jump process to BM:

$$dS_t = S_t \sigma dW_t + J_t.$$

where J_t represents the jump process, and for which we considered two cases:

- 1 uniformly distributed jump of a random size (2-5% change in asset price);
- still 1 jump, but of a greater magnitude (6-9% change in asset price).

The latter extreme jumps also deserve attention, while it will let us stress test our estimators, plus empirical studies prove possibility of such price movements. For instance in the papers Bakshi et al. (2008) or Malkiel et al. (2009) authors mention that there were 69 days in the history on which the stock market has dropped by more than 5%. As a stock market they consider DJIA index, which means for every single stock probability of this event is even higher.

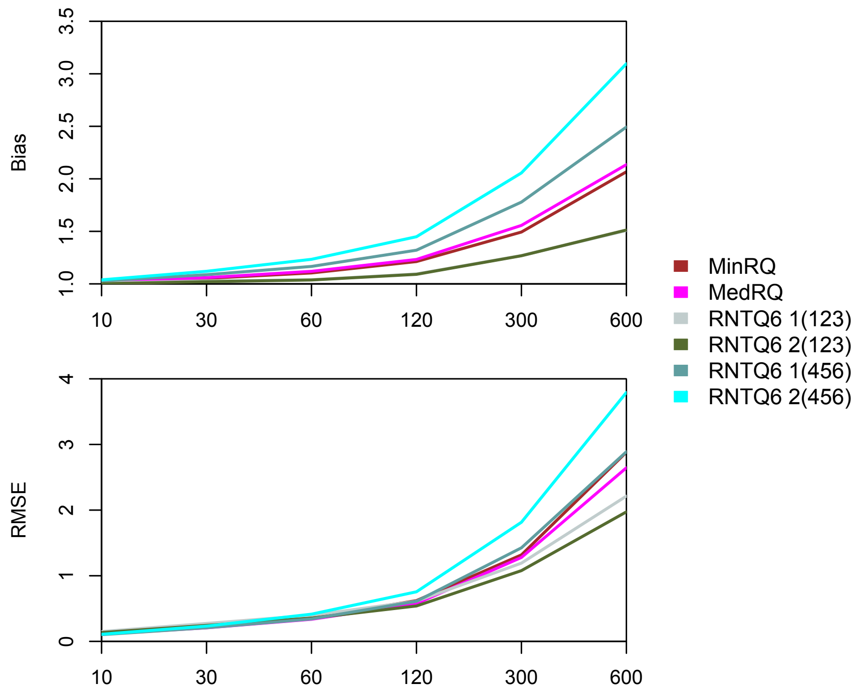

Figure 4.2: Brownian motion simulation with 1 jump of a randomly distributed 2-5% size

Results of these jump simulations are presented at the Figures 4.2 and 4.3. We have excluded estimators `RQ`, `TQ`, `MPQ4` and `MPQ5` from graphical comparison due to their pure performance, however detailed overview is given in Table 4.2 and Table in Appendix.

Patterns of biases and RMSE values for all the estimators are similar between two cases we have considered, however, in the situation with bigger jumps, we observe somehow lower values: biases are lower in round 1.8-2.5 times, while RMSE in 1.4-2.0 times. The difference grows proportionally to the growing pre-averaging window size.

Another interesting fact about `RNTQ6` estimators was observed. Andersen et al. (2011) suggested, that it is useless to apply lower order statistics of underlying returns in `RNT` esti-

			BIAS		
	RQ	TQ	MPQ4	MPQ5	MinRQ
10	0,99942	0,99876	0,99854	0,99887	0,99967
30	0,99636	1,00055	1,00079	1,00035	0,99761
60	0,99535	0,99724	0,99585	0,99492	0,99643
120	0,99614	0,99695	0,99545	0,99437	0,99586
300	1,01007	0,99632	0,99661	0,99739	0,99709
600	1,01396	1,00736	1,01093	1,01573	0,99928
	MedRQ	RNTQ6 1(123)	RNTQ6 2(123)	RNTQ6 1(456)	RNTQ6 2(456)
10	0,99976	0,99960	0,99457	1,00082	1,00013
30	0,99728	1,00236	0,99671	0,99898	0,99842
60	0,99619	0,99091	0,98718	0,99887	0,99799
120	0,99521	0,99137	0,98702	0,99720	0,99703
300	0,99918	0,99893	0,98916	0,99902	1,00040
600	1,00029	1,04247	1,03274	1,00749	1,00537
			RMSE		
	RQ	TQ	MPQ4	MPQ5	MinRQ
10	0,06596	0,07330	0,07527	0,07701	0,08207
30	0,11286	0,12718	0,13083	0,13235	0,14341
60	0,16128	0,18147	0,18618	0,18891	0,20278
120	0,23371	0,25581	0,26518	0,26863	0,28224
300	0,38322	0,40906	0,42531	0,43825	0,45107
600	0,53576	0,62321	0,66411	0,69786	0,66963
	MedRQ	RNTQ6 1(123)	RNTQ6 2(123)	RNTQ6 1(456)	RNTQ6 2(456)
10	0,07324	0,10845	0,09737	0,06556	0,06442
30	0,12812	0,18334	0,16645	0,11329	0,11094
60	0,18119	0,25978	0,23324	0,16388	0,15942
120	0,24968	0,37128	0,33598	0,22972	0,22407
300	0,40905	0,61484	0,55355	0,37815	0,37301
600	0,60263	1,07736	0,97339	0,57090	0,55436

Table 4.1: Brownian motion asset price simulation

	Bias				
	RQ	TQ	MPQ4	MPQ5	MinRQ
10 sec	28272,04	2,51465	1,48904	1,25536	1,01606
30 sec	9362,887	3,22015	1,85813	1,49889	1,05211
60 sec	4680,045	3,79553	2,17393	1,72069	1,10397
120 sec	2349,351	4,34489	2,61002	2,05687	1,21231
300 sec	948,3998	5,53123	3,51648	2,79007	1,49342
600 sec	484,0827	6,72046	4,49893	3,64871	2,06711
	MedRQ	RNTQ6 1(123)	RNTQ6 2(123)	RNTQ6 1(456)	RNTQ6 2(456)
10 sec	1,01781	1,00826	1,00148	1,02773	1,03838
30 sec	1,05907	1,02688	1,02051	1,08712	1,11961
60 sec	1,11925	1,03906	1,03693	1,16598	1,23361
120 sec	1,23226	1,09445	1,09072	1,32066	1,44867
300 sec	1,55646	1,26897	1,26764	1,77861	2,05633
600 sec	2,13526	1,50751	1,51277	2,49325	3,09623
	RMSE				
	RQ	TQ	MPQ4	MPQ5	MinRQ
10 sec	37077,67	2,15031	0,68935	0,36148	0,12413
30 sec	12295,42	3,16764	1,21572	0,71341	0,22185
60 sec	6163,644	3,98274	1,67833	1,05269	0,36142
120 sec	3108,752	4,91599	2,39031	1,58756	0,62124
300 sec	1281,055	7,02737	3,89827	2,83822	1,31755
600 sec	678,3788	9,62314	6,14168	4,86702	2,88209
	MedRQ	RNTQ6 1(123)	RNTQ6 2(123)	RNTQ6 1(456)	RNTQ6 2(456)
10 sec	0,11072	0,14967	0,13451	0,10213	0,10788
30 sec	0,20621	0,27357	0,24372	0,20758	0,23274
60 sec	0,33655	0,39809	0,35362	0,34217	0,41428
120 sec	0,57601	0,61173	0,53928	0,61088	0,75551
300 sec	1,27862	1,19208	1,07784	1,42551	1,81474
600 sec	2,64457	2,21311	1,97303	2,88913	3,79192

Table 4.2: Brownian motion simulation with 1 jump of a randomly distributed 2-5% size

mators, since they are relatively more affected by market microstructure noise. On the contrary to that, constructed by us estimators `RNT6 1(123)` and `RNT6 2(123)` outperformed significantly all the other estimators. Finishing with almost identical results, during these simulations they appeared to be way more efficient than the estimators `RNT6 1(456)` and `RNT6 2(456)`, which use higher order statistics returns. This circumstance motivated us to pay more attention to `LOS RNTQ` estimators and run analogous simulations specifically for a group of these estimators (Section 4.4).

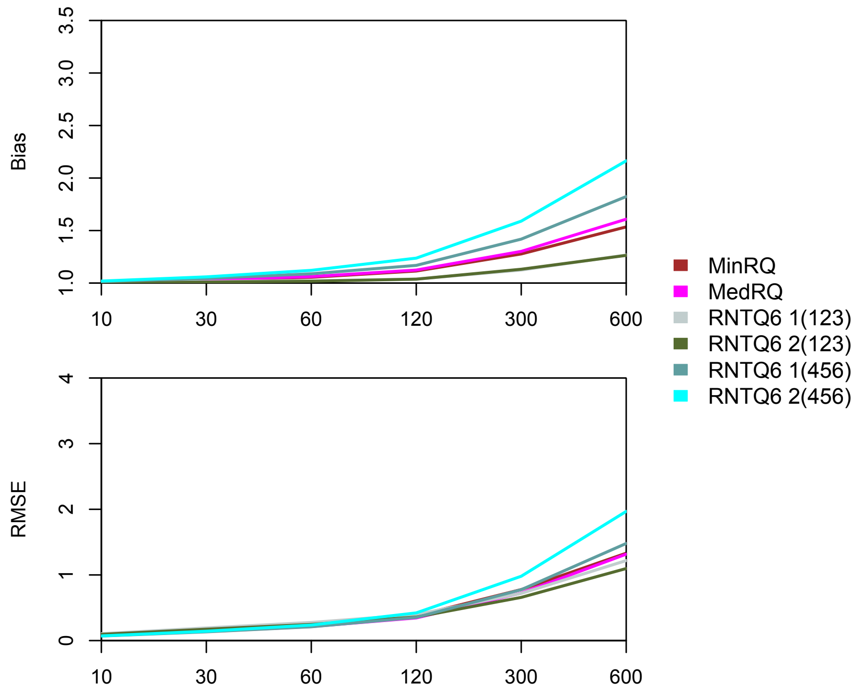

Figure 4.3: Brownian motion simulation with 1 jump of a randomly distributed 6-9% size

4.2 Stochastic volatility model

Another simulations we have performed were based on the model discussed in the paper Andersen et al. (2009). Stochastic volatility model with intraday U-shape volatility pattern is described there by:

$$\begin{aligned}
dS(t) &= \sigma_u(t)\sigma_{sv}(t)dW(t) \\
\sigma_{sv}^2(t) &= \sigma_1^2(t) + \sigma_2^2(t) \\
d\sigma_1^2(t) &= k_1\left[\theta_1 - \sigma_1^2(t)\right]dt + \eta_1\sigma_1(t)dW_1(t) \\
d\sigma_2^2(t) &= k_2\left[\theta_2 - \sigma_2^2(t)\right]dt + \eta_2\sigma_2(t)dW_2(t) \\
\sigma_u(t) &= C + Ae^{-at} + Be^{-b(1-t)}, \quad t \in [0,1].
\end{aligned}$$

The set of parameters is taken as:

$$\begin{aligned}
k_1 &= 0.6, & \theta_1 &= 1.0582, & \eta_1 &= 0.2, & \rho_1 &= 0.9, \\
k_2 &= 0.1, & \theta_2 &= 0.5291, & \eta_2 &= 0.1, & \rho_1 &= -0.4.
\end{aligned}$$

Last pair of coefficients ρ_1, ρ_2 defines the instantaneous correlations $\rho_1 = corr(dW(t), dW_1(t))$ and $\rho_2 = corr(dW(t), dW_2(t))$, while the processes W_1 and W_2 are independent.

Equation for $\sigma_u(t)$, taken with parameters $A = 0.75$, $B = 0.25$, $C = 0.88929198$, $a = 10$ and $b = 10$, gives us asymmetric U-shaped intraday variance curve. At the beginning of the trading day the variance is more than 3 times bigger than midday variance, while at the close it is around 1.5 times the midday value (Figure 4.4).

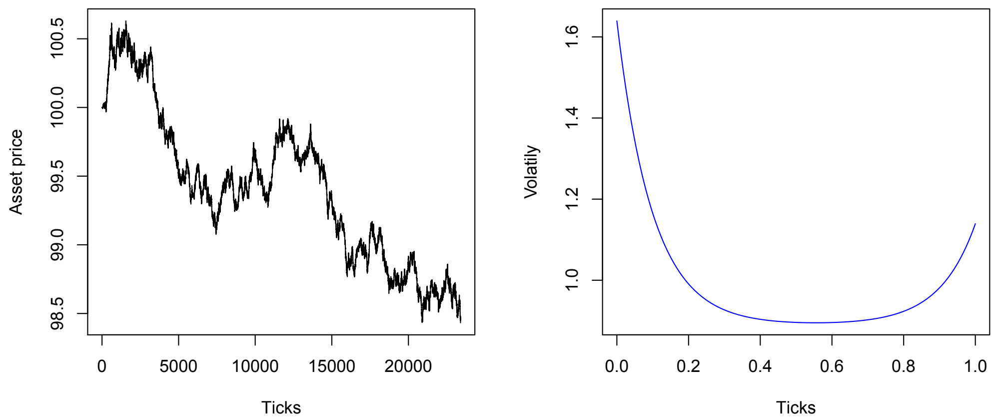

Figure 4.4: U-shaped intraday volatility

Such volatility shape is an empirical observation that was studied, for instance, by Hong and Wang (2000). They have come up with several conclusions that mean and volatility of returns over trading periods have a U-shaped pattern, as well as around the market opening and closing times trading activity is higher. Increased activity in the morning can be explained by the willingness of risk averse investors to hedge their assets, while at the end of the day informed speculators try to open profitable positions.

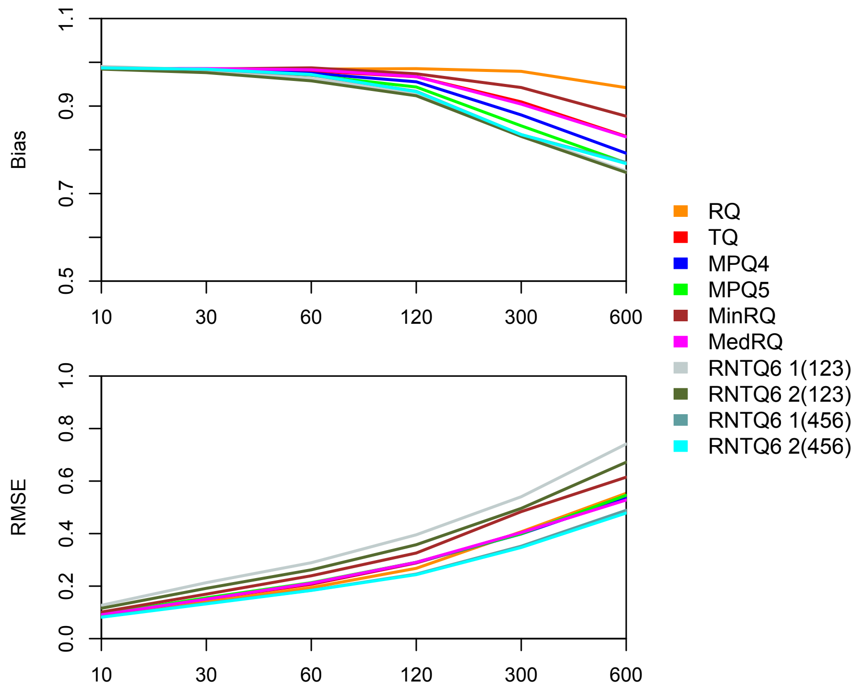

Figure 4.5: Stochastic volatility model with intraday U-shape

The SV-U model sampled on an equispaced time grid allows to isolate possible finite sample biases of the considered estimators due to time variation in volatility. According to Andersen et al. (2010), the effect of the applied U-shape is to make neighboring returns heterogeneous which tends to produce a downward bias in all the estimators.

Results, we have received, are completely in line with those, that authors present in the referenced paper. Thus, all estimators showed similar downward bias. Most efficient came out to be `RQ` and `MinRQ` estimators, as those that work with least quantity of adjacent returns, namely two. Quite close to them are `TQ` and `MedRQ` with almost identical results in bias and RMSE values (involves three adjacent returns). After that come `MPQ4`, `MPQ5` and the least efficient group `RNTQ6 1(123)`, `RNTQ6 2(123)`, `RNTQ6 1(456)`, `RNTQ6 2(456)` as the least "local" estimators which capture 6 neighboring returns. Comparing to the Brownian motion simulations, in case of SV-U model `RNTQ6 1(123)` and `RNTQ6 2(123)` had the poorest performance out of all estimators, finishing with the largest RMSE error.

	Bias				
	RQ	TQ	MPQ4	MPQ5	MinRQ
10 sec	0,98735	0,98728	0,98736	0,98761	0,98546
30 sec	0,98366	0,98424	0,98423	0,98333	0,98473
60 sec	0,98434	0,98029	0,97502	0,97055	0,98733
120 sec	0,98549	0,96705	0,95545	0,94351	0,97375
300 sec	0,97934	0,90957	0,87970	0,85484	0,94240
600 sec	0,94201	0,83068	0,79213	0,77008	0,87683
	MedRQ	RNTQ6 1(123)	RNTQ6 2(123)	RNTQ6 1(456)	RNTQ6 2(456)
10 sec	0,98591	0,99058	0,98437	0,98752	0,98702
30 sec	0,98569	0,98286	0,97640	0,98394	0,98428
60 sec	0,98297	0,96310	0,95763	0,97095	0,97226
120 sec	0,96724	0,92929	0,92353	0,93381	0,93290
300 sec	0,90444	0,83480	0,83062	0,83457	0,83491
600 sec	0,82954	0,75129	0,74823	0,77117	0,76845
	RMSE				
	RQ	TQ	MPQ4	MPQ5	MinRQ
10 sec	0,08507	0,09030	0,09248	0,09351	0,10121
30 sec	0,14012	0,15090	0,15403	0,15517	0,16945
60 sec	0,19441	0,20705	0,21069	0,21294	0,23928
120 sec	0,26773	0,28837	0,29000	0,29081	0,32600
300 sec	0,40692	0,40391	0,39882	0,39963	0,48366
600 sec	0,55343	0,52896	0,53582	0,54587	0,61481
	MedRQ	RNTQ6 1(123)	RNTQ6 2(123)	RNTQ6 1(456)	RNTQ6 2(456)
10 sec	0,09090	0,12695	0,11562	0,08256	0,08143
30 sec	0,15039	0,21315	0,19149	0,13413	0,13231
60 sec	0,21104	0,28904	0,26197	0,18423	0,18345
120 sec	0,29065	0,39546	0,35762	0,24573	0,24362
300 sec	0,40388	0,54041	0,49608	0,35203	0,34705
600 sec	0,52788	0,74117	0,67181	0,48855	0,47822

Table 4.3: Simulation of stochastic volatility model with intraday U-shape

4.3 Brownian motion with sparse sampling

Data irregularity is one of the important issues of financial data. Quotes arrivals and stock price movements happen not on a regular basis with some fixed time period, but usually randomly to some extent, making lags of different size between time points.

Conducting simulations with sparse sampling can be helpful at investigating influence of such market data structure on IQ estimators. Among papers that have already described some results of such simulations are applied to volatility or quarticity estimations are Zhang et al. (2005), Andersen et al. (2011).

Initially, for each trading day we were generating standard Brownian motion process with 23400 values. At the next step, values out of this time series were picked using Poisson distribution with $\lambda = 2$, in order to get non-homogeneous data time-arrivals. This approach was providing us with a sample, whose size varied on average between 10850 and 11050 time points.

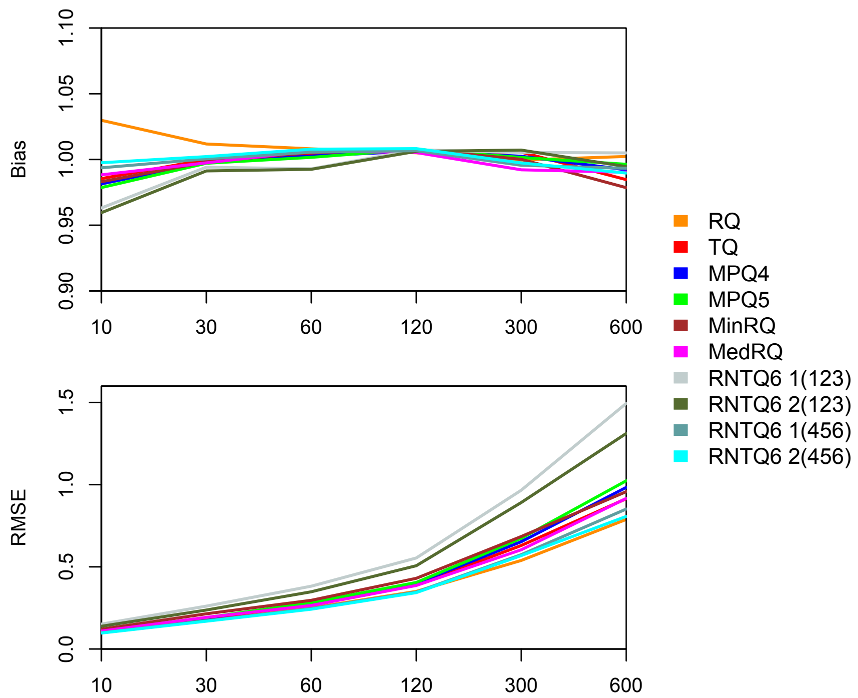

Figure 4.6: Brownian motion simulation with sparse sampling

Detailed results of this simulation are presented in Table 4.4, as well as illustrated in Figure 4.6. Firstly, the scale of Bias and RMSE values is comparable to the one that BM and SV-U models had: all the estimators' biases vary in a range 0.96-1.04 and RMSE around 0-1.5. Simulations with random jumps demonstrated considerably greater values, then sparse sampling case.

Picking small 10 second sampling window resulted in higher biases of `RNTQ6 1(123)`, `RNTQ6 2(123)` and `RQ` estimators: around 0.96 for the first two, and 1.03 for the latter one. At the very same moment, RMSE stays lowest - 0.11-0.12. All the other estimators' find themselves in the range 0.98-1.00.

Increasing the sampling window size up to 120 seconds results in narrower range of biases for all of the estimators pushing them closer to 1.00. Further less frequent sampling leads to some overall downward bias. Parallel to that, RMSE rises at a constant pace up to values 1-1.5.

Within this particular simulation, `RNTQ6 1(123)` and `RNTQ6 2(123)` performed worth then all the other estimators, while having highest downward biases all way long till 120 sec window, plus demonstrating distinctly higher RMSE error.

			Bias		
	RQ	TQ	MPQ4	MPQ5	MinRQ
10 sec	1,02980	0,98546	0,98141	0,97864	0,98310
30 sec	1,01179	0,99996	0,99805	0,99712	0,99807
60 sec	1,00810	1,00447	1,00261	1,00162	1,00549
120 sec	1,00509	1,00637	1,00636	1,00778	1,00772
300 sec	0,99957	1,00558	1,00246	1,00153	1,00014
600 sec	1,00230	0,98470	0,99216	0,99636	0,97855
	MedRQ	RNTQ6 1(123)	RNTQ6 2(123)	RNTQ6 1(456)	RNTQ6 2(456)
10 sec	0,98833	0,96311	0,95952	0,99368	0,99751
30 sec	0,99744	0,99413	0,99129	1,00047	1,00220
60 sec	1,00660	0,99287	0,99253	1,00557	1,00783
120 sec	1,00527	1,00738	1,00617	1,00670	1,00827
300 sec	0,99214	1,00535	1,00715	0,99559	0,99738
600 sec	0,99030	1,00501	0,99459	0,99329	0,98967
			RMSE		
	RQ	TQ	MPQ4	MPQ5	MinRQ
10 sec	0,11445	0,10644	0,10921	0,11056	0,12037
30 sec	0,17762	0,18528	0,18872	0,19067	0,21416
60 sec	0,24978	0,27115	0,27772	0,28236	0,29624
120 sec	0,35041	0,39002	0,39933	0,40479	0,42994
300 sec	0,53870	0,62971	0,65286	0,67301	0,68426
600 sec	0,78821	0,91413	0,98298	1,02337	0,95737
	MedRQ	RNTQ6 1(123)	RNTQ6 2(123)	RNTQ6 1(456)	RNTQ6 2(456)
10 sec	0,10897	0,15132	0,13812	0,09863	0,09694
30 sec	0,19239	0,26081	0,23702	0,17117	0,16901
60 sec	0,26477	0,38223	0,34810	0,24596	0,24208
120 sec	0,38654	0,55367	0,50658	0,34660	0,34288
300 sec	0,60370	0,96610	0,89120	0,57402	0,56829
600 sec	0,91606	1,49427	1,31144	0,85148	0,80648

Table 4.4: Brownian motion with sparse sampling

4.4 Lower order statistics RNT quarticity estimators

At the very beginning of our research we have decided to examine `RNTQ6` with both low and high order statistics. As was mentioned already in the Section 4.1, authors Andersen et al. (2011) made an assumption that `LOS RNTQ` estimators are more affected by market microstructure noise and did not include them to overall simulations analysis. Without any verification that decision seemed to us a little bit premature, while based on the simulations performed in Section 4.1, estimators `RNTQ6 1(123)`, `RNTQ6 2(123)` were the best, in terms of bias and RMSE error, under the jump presence and demonstrated in general decent performance in simulations with stochastic volatility and sparse sampling of stock returns.

We have examined a group of `RNT` estimators which covered various order statistics configurations:

- `RNTQ5 1(123)  RNTQ5 2(123);`
- `RNTQ5 1(345)  RNTQ5 2(345);`
- `RNTQ6 1(123)  RNTQ6 2(123);`
- `RNTQ6 1(456)  RNTQ6 2(456);`
- `RNTQ7 1(1234)  RNTQ7 2(1234);`
- `RNTQ7 1(4567).`

In proposed setup estimator `RNTQ7 2(4567)` was omitted due to pure efficiency caused by low jump robustness.

Guided by formulas stated in the Section 2.3, on addition to scaling coefficients of `RNTQ6` estimator (Table 2.1), we have calculated analogous values for `RNTQ5` and `RNTQ7` (Table 4.5).

	$d_{(1,3)}(4)$	$d_{(2,3)}(4)$	$d_{(3,3)}(4)$		$d_{(j,\boldsymbol{I})}(4)$
$RNTQ51(123)$	35,14029	5,75253	1,44264		3,67611
$RNTQ52(123)$	35,03199	5,74508	1,44339		1,31886
$RNTQ51(345)$	1,44314	0,39879	0,08642		2,60658
$RNTQ52(345)$	1,43542	0,39656	0,08632		1,21894
	$d_{(1,4)}(4)$	$d_{(2,4)}(4)$	$d_{(3,4)}(4)$	$d_{(4,4)}(4)$	$d_{(j,\boldsymbol{I})}(4)$
$RNTQ71(1234)$	104,37888	18,57741	5,31021	1,85594	4,56927
$RNTQ74(1234)$	105,33341	18,71883	5,29974	1,84848	0,44952
$RNTQ71(4567)$	1,85712	0,69869	0,25216	0,06739	2,70389

Table 4.5: Scaling factors of RNTQ5 and RNTQ7 estimators for different order statistics

One can observe, that together with the rise of the returns quantity, coefficients grow even more, with a sharp distinction between the groups of lower order and higher order returns.

Recalling Equation 2.24 from the Section 2.3 we write down expression for the asymptotic distribution of `RNTQ` estimator for pure BM process without jumps:

$$\sqrt{n}\left(RNTQ_N^{(j,\boldsymbol{I})} - \int_0^1 \sigma_s^4 ds\right) \xrightarrow{\mathscr{L}} \mathrm{N}\left(0, \eta(j,\boldsymbol{I};4)\int_0^1 \sigma_s^8 ds\right), \quad j = 1,\ldots,H. \tag{4.1}$$

While trying to approximate to some extent the efficiency factors $\eta(j, \boldsymbol{I}; 4)$ of our target group estimators, we have received values postulated in the Table 5 in Appendix. Analogously to the `MPV` estimator's property mentioned in Andersen et al. (2011), scrutinized `RNTQ` estimators, under the no-jump null hypothesis, have a tendency to improve efficiency when block size of returns gets smaller.

Another important result is, that under pure BM process, `HOS RNTQ` perform definitely better then `LOS RNTQ`. Estimators `RNTQ5 1(345)`, `RNTQ 2(345)`, `RNTQ6 1(456)`, `RNT6 2(456)` and even `RNT7 1(4567)` have asymptotic variances settled around values 10-11. Meanwhile, `LOS RNTQ` estimators starting from `RNTQ5 1(123)` constantly grow in variance measure, hitting values up to 30-40. These evidence are quite in line with the simulation results received in Section 4.1.

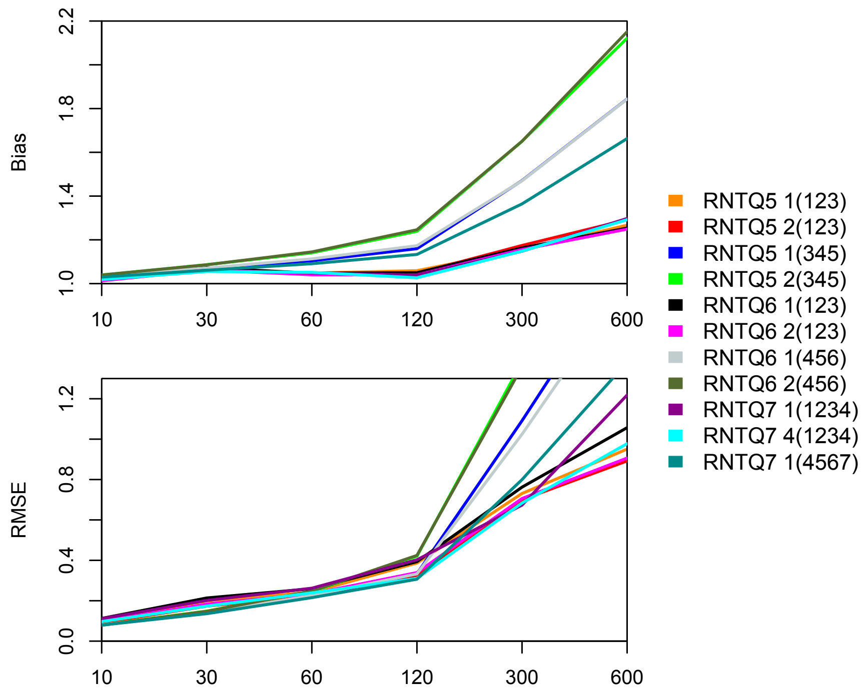

Figure 4.7: RNT quarticity estimators applied to BM stochastic process with 1 jump of a randomly distributed 2-5% size

Similar to the simulations already covered in the current chapter, we have examined `RNTQ` estimators with appearance of random jumps, SV-U model and sparse sampling. We have used the same models and simulation parameters as before.

BM model with one random jump clearly showed significantly greater biases of estimators `RNTQ5 1(345)`, `RNTQ5 2(345)`, `RNTQ6 1(456)`, `RNTQ6 2(456)` and `RNTQ7 4(4567)` (especially with a sampling window great then 120 seconds). All the other estimators, while grouped together quite tightly, together show relatively small bias (Figure 4.7). With RMSE errors situation looks quite similar, with a breaking point again in 120 seconds. We can definitely say that `LOS RNTQ` are more robust to the presence of a random jump within trading interval. This seems reasonable, while picking values out of group of lower order returns, most surely will let us omit the jump component, in case such is present withing observable interval. This simulation does not demonstrate difference between, say, estimators `RNTQ5 1(345)` and `RNTQ7 1(4567)`, but we suppose it will be more evident under presence of greater quantity of jumps, which can be verified separately.

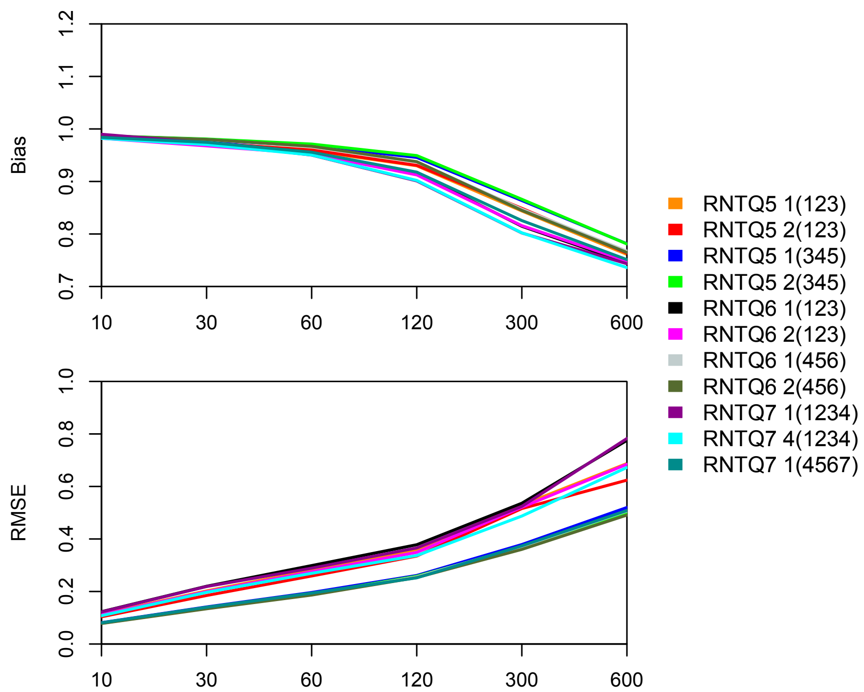

Figure 4.8: RNT quarticity estimators applied to stochastic volatility model with intraday U-shape

Simulation of stochastic volatility in this case supported previous results, obtained in the Section 4.2. All estimators tend to have downward bias, and while closely grouped, it is hard to single out some particular one, significantly better then the others (Figure 4.8). Estimators like `RNTQ5 2(345)` or `RNTQ6 2(456)` are slightly more efficient, both in terms of bias and RMSE error. Overall we can say that applied to SV-U model, `HOS RNTQ` estimators are a bit more efficient then `LOS RNTQ`.

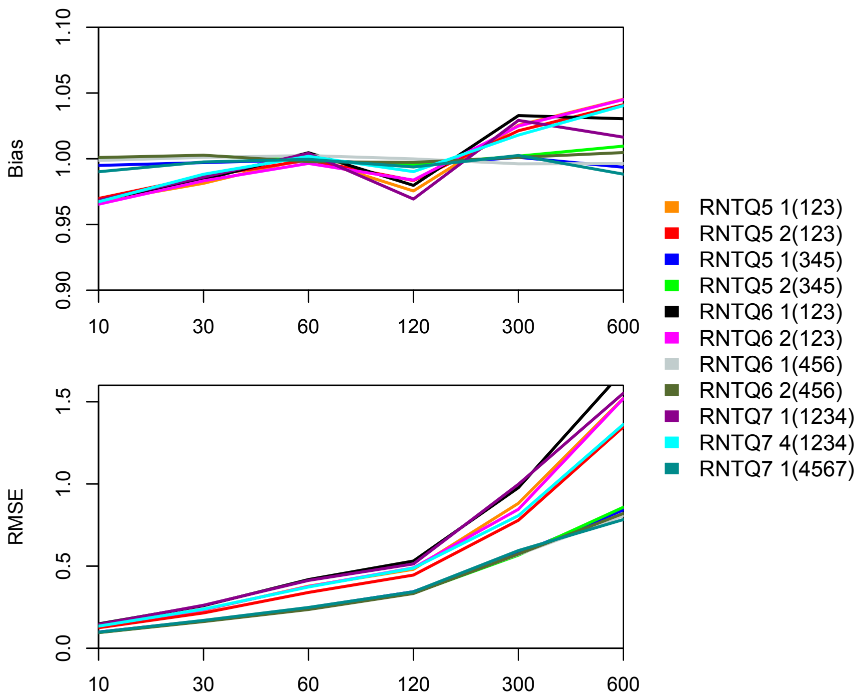

Figure 4.9: RNT quarticity estimators applied to BM stochastic process with sparse sampling

Last simulation showed instability of `LOS RNTQ` estimators against sampling window size. Figure 4.9 reveals that choice of sampling window is quite important - picking appropriate one can let us reach lower levels of bias. In general, based on Figure 4.9 and Figure 4.6 choosing sampling windows 10-30 seconds and less, can lead to rise in bias. On the contrary to that, `HOS RNTQ` estimators revealed constantly good performance, all the time demonstrating low bias. Coupled with lower RMSE values, in case we speak about non-equidistant returns, `HOS RNTQ` seem to be more attractive then `LOS RNTQ`.

Conducted in the Sections 4.1-4.3 simulations provided us with results that are correlating with those, presented by Andersen et al. (2010) and Andersen et al. (2011). This reassures to some extent, that created computational setup of this work is reasonable and adequate, which let us conclude, that results of the Section 4.4 are reliable enough and should not be rejected.

Conclusions

Proposed work was intended to provide sufficient analysis of several types of integrated quarticity estimators, that have been discussed in the recent literature. A variety of papers has been considered, among which fundamental role played Barndorff-Nielsen and Shephard (2004), Andersen et al. (2009), Andersen et al. (2010) as well as Andersen et al. (2011). Theoretical approaches and performed simulations, stated in this work, were to certain extent replicating models used in these papers, thus one of our main goals was to extend them and reveal issues that were left out by previous authors. In this context we have payed special attention to the Robust Neighborhood Truncation Estimators with lower order statistics log-returns.

Empirical studies were focused on calculation of integrated quarticity for stocks of three companies *Susquehanna Bancshares*, *Pfizer* and *Exxon Mobil.* While they were characterized by different market capitalizations and trading amounts, we saw a significant influence of these parameters on log-returns magnitude along the whole time period. *Susquehanna Bancshares* had, on average, the greatest returns while *Exxon Mobil* the smallest ones, which eventually influenced scales of integrated quarticities. If the price changes were not too rough, estimators `MPV(4,4)` and `MedRQ` were softening the jumps better then `MPV(3,4)` and `MinRQ`, respectively. Estimators `RNTQ6 1(123)` and `RNTQ6 1(456)` were producing much lower values then all the other measures, suppressing most of the stocks' volatility. Among them, `RNTQ6 1(123)` gave even smother results then `RNTQ6 1(456)`, which was the first hint at the greater jump robustness of `RNT` estimators which work with lower order statistics log-returns.

Brownian motion simulation, as it was expected, showed almost no bias behind each of the estimators. Only `RNTQ6 1(123)` and `RNTQ6 2(123)` had some small divergences at the sampling window sizes above 30 seconds. Meanwhile, in the second BM model that introduced random jumps of different sizes, we have witnessed superiority of estimators `RNTQ6 1(123)` and `RNTQ6 1(123)`, both in terms of bias and RMSE error, above all the other estimators, including `RNTQ6 1(456)` and `RNTQ6 2(456)`. During the simulations of stochastic volatility model, `RNTQ6`, together with `MPQ4` and `MPQ5` estimators, performed not as good as `MinRQ`, `MedRQ` and demonstrated the greatest downward bias. Within this test, estimators with lower quantity of included adjacent returns definitely are more efficient.

Simulation with sparse sampling did not let us to distinguish estimators that much. All of them had considerably low bias, however right choice of sampling window did seem to have greater influence. Thus, high frequencies like 10 seconds, as well as as the ones above 600 seconds were provoking higher bias.

After confronting `RNT6` with other multipower variation estimators and nearest neighbor truncation estimators, we were interested in direct comparison of different combinations of `LOS RNTQ` and `HOS RNTQ` estimators. With this purpose we considered `RNTQ5`, `RNTQ6` and `RNTQ7` estimators. After derivation of the asymptotic covariance matrix of their joint distribution under the condition of BM process and no-jump hypothesis, it became evident that `HOS RNTQ` is more efficient then `LOS RNTQ`, which was not the case in all further simulations. Thus, `LOS RNTQ` estimators were much more jump robust then `HOS RNTQ`, and they also showed decent performance in stochastic volatility model and Brownian motion with sparse sampling simulations. While all the previous simulation results are in line with respective literature, derived efficiency of `LOS RNTQ` estimators appears to be reliable enough and should not be rejected.

Possible way to extend this research include examination of bigger set of more diverse `RNTQ` estimators and their assessment with simulations that would mix several price process models in one time.

Appendix

			Bias		
	RQ	TQ	MPQ4	MPQ5	MinRQ
10 sec	786553,6	3,98603	1,69273	1,30078	1,00733
30 sec	260881,8	5,39626	2,21141	1,58229	1,02738
60 sec	130552,6	6,66533	2,74578	1,88623	1,05471
120 sec	65277,85	7,79578	3,34478	2,27526	1,11457
300 sec	26122,99	10,15922	4,68741	3,19624	1,27717
600 sec	13061,74	12,72215	6,21651	4,32148	1,53474
	MedRQ	RNTQ6 1(123)	RNTQ6 2(123)	RNTQ6 1(456)	RNTQ6 2(456)
10 sec	1,00838	1,00412	1,00186	1,01446	1,01894
30 sec	1,02964	1,01364	1,00902	1,04347	1,05897
60 sec	1,06142	1,02173	1,01714	1,08805	1,12096
120 sec	1,12372	1,03798	1,03711	1,16816	1,23721
300 sec	1,30165	1,12363	1,13161	1,41818	1,58942
600 sec	1,60811	1,26709	1,26337	1,82332	2,16362
			RMSE		
	RQ	TQ	MPQ4	MPQ5	MinRQ
10 sec	937487,8	4,07298	0,93653	0,40484	0,08568
30 sec	311093,3	5,94871	1,63982	0,78596	0,15341
60 sec	155857,2	7,94352	2,42041	1,23014	0,23106
120 sec	78127,25	9,75471	3,31651	1,79659	0,37733
300 sec	31521,12	13,15809	5,30488	3,18718	0,77717
600 sec	15985,63	16,88308	7,61757	4,94321	1,33138
	MedRQ	RNTQ6 1(123)	RNTQ6 2(123)	RNTQ6 1(456)	RNTQ6 2(456)
10 sec	0,07559	0,10513	0,09444	0,06913	0,07027
30 sec	0,13788	0,19225	0,17275	0,13292	0,14206
60 sec	0,21335	0,27619	0,24612	0,20993	0,23647
120 sec	0,34517	0,39654	0,35869	0,35367	0,42177
300 sec	0,72315	0,71956	0,65767	0,77518	0,97992
600 sec	1,31658	1,21861	1,09741	1,47945	1,96791

Table 1: Brownian motion simulation with 1 extreme jump (5-9% of the stock price)

			Bias			
	RNTQ5 1(123)	RNTQ5 2(123)	RNTQ5 1(345)	RNTQ5 2(345)	RNTQ6 1(123)	RNTQ6 2(123)
10 sec	1,01595	1,01714	1,03251	1,03902	1,01803	1,01277
30 sec	1,06884	1,06384	1,06986	1,08644	1,07214	1,06166
60 sec	1,04828	1,04243	1,10415	1,14091	1,04773	1,04036
120 sec	1,05894	1,04686	1,15961	1,23902	1,04978	1,04007
300 sec	1,15480	1,17421	1,47079	1,65042	1,16321	1,15585
600 sec	1,26718	1,29200	1,84380	2,12166	1,25366	1,24884
	RNTQ6 1(456)	RNTQ6 2(456)	RNTQ7 1(1234)	RNTQ7 4(1234)	RNTQ7 1(4567)	
10 sec	1,03537	1,03938	1,02424	1,01876	1,02974	
30 sec	1,07072	1,08602	1,06103	1,05535	1,06106	
60 sec	1,11296	1,14486	1,04859	1,05143	1,09094	
120 sec	1,17314	1,24609	1,03950	1,02666	1,13329	
300 sec	1,46986	1,65052	1,14797	1,14839	1,36425	
600 sec	1,84329	2,15116	1,29759	1,29187	1,66290	
			RMSE			
	RNTQ5 1(123)	RNTQ5 2(123)	RNTQ5 1(345)	RNTQ5 2(345)	RNTQ6 1(123)	RNTQ6 2(123)
10 sec	0,10131	0,09078	0,08063	0,08014	0,11239	0,09803
30 sec	0,19870	0,17308	0,14218	0,14879	0,21367	0,18321
60 sec	0,24277	0,22526	0,22399	0,24604	0,25952	0,23462
120 sec	0,38656	0,32513	0,33293	0,42041	0,39342	0,33933
300 sec	0,73090	0,70030	1,09130	1,38221	0,76229	0,70603
600 sec	0,95067	0,89221	1,88819	2,01343	1,05688	0,90631
	RNTQ6 1(456)	RNTQ6 2(456)	RNTQ7 1(1234)	RNTQ7 4(1234)	RNTQ7 1(4567)	
10 sec	0,08038	0,07867	0,11126	0,09670	0,07912	
30 sec	0,14196	0,14741	0,20148	0,17324	0,13589	
60 sec	0,22553	0,24773	0,26212	0,23889	0,21527	
120 sec	0,33273	0,42431	0,40249	0,31054	0,30648	
300 sec	1,02427	1,35861	0,67353	0,68373	0,79944	
600 sec	1,77769	2,06567	1,21832	0,97761	1,38073	

Table 2: RNTQ estimators performance at BM process simulation with 1 jump of a randomly distributed 2-5% size

			Bias			
	RNTQ5 1(123)	RNTQ5 2(123)	RNTQ5 1(345)	RNTQ5 2(345)	RNTQ6 1(123)	RNTQ6 2(123)
10 sec	0,98269	0,98310	0,98444	0,98712	0,98790	0,98220
30 sec	0,97358	0,97324	0,97830	0,98091	0,97454	0,96779
60 sec	0,95832	0,96001	0,96793	0,97121	0,95495	0,95251
120 sec	0,92987	0,93069	0,94580	0,94926	0,91584	0,91247
300 sec	0,84383	0,84882	0,86368	0,86606	0,81469	0,81589
600 sec	0,76111	0,76357	0,78139	0,78096	0,74565	0,74852
	RNTQ6 1(456)	RNTQ6 2(456)	RNTQ7 1(1234)	RNTQ7 4(1234)	RNTQ7 1(4567)	
10 sec	0,98679	0,98679	0,98988	0,98211	0,98388	
30 sec	0,97987	0,98011	0,97347	0,97026	0,97456	
60 sec	0,96646	0,96695	0,95020	0,95045	0,95570	
120 sec	0,93745	0,93761	0,90100	0,90200	0,91777	
300 sec	0,84725	0,84500	0,80190	0,80218	0,82550	
600 sec	0,76732	0,76470	0,74338	0,73601	0,75117	
			RMSE			
	RNTQ5 1(123)	RNTQ5 2(123)	RNTQ5 1(345)	RNTQ5 2(345)	RNTQ6 1(123)	RNTQ6 2(123)
10 sec	0,11361	0,10368	0,08129	0,07907	0,12313	0,11118
30 sec	0,20229	0,18475	0,14118	0,13599	0,21994	0,19956
60 sec	0,28083	0,25904	0,19619	0,19079	0,29868	0,27520
120 sec	0,35980	0,33506	0,26105	0,25816	0,37813	0,34900
300 sec	0,52923	0,51588	0,37884	0,36788	0,53614	0,52099
600 sec	0,68734	0,62407	0,51987	0,49809	0,77555	0,68467
	RNTQ6 1(456)	RNTQ6 2(456)	RNTQ7 1(1234)	RNTQ7 4(1234)	RNTQ7 1(4567)	
10 sec	0,07978	0,07810	0,12220	0,10776	0,08073	
30 sec	0,13867	0,13415	0,21887	0,19683	0,13998	
60 sec	0,18968	0,18642	0,28889	0,26920	0,19392	
120 sec	0,25561	0,25311	0,36579	0,33697	0,25210	
300 sec	0,37146	0,36037	0,52421	0,48731	0,37381	
600 sec	0,50625	0,49111	0,78267	0,67229	0,50815	

Table 3: RNTQ estimators performance at SV-U simulation

	Bias					
	RNTQ5 1(123)	RNTQ5 2(123)	RNTQ5 1(345)	RNTQ5 2(345)	RNTQ6 1(123)	RNTQ6 2(123)
10 sec	0,96739	0,96978	0,99492	0,99968	0,96833	0,96541
30 sec	0,98133	0,98651	0,99706	1,00234	0,98423	0,98323
60 sec	1,00067	0,99895	0,99923	0,99869	1,00462	0,99644
120 sec	0,97550	0,98343	0,99512	0,99574	0,97974	0,98377
300 sec	1,02549	1,02132	1,00110	1,00198	1,03284	1,02489
600 sec	1,04530	1,04113	0,99355	1,00961	1,03044	1,04501
	RNTQ6 1(456)	RNTQ6 2(456)	RNTQ7 1(1234)	RNTQ7 4(1234)	RNTQ7 1(4567)	
10 sec	0,99869	1,00094	0,96844	0,96727	0,99011	
30 sec	1,00107	1,00269	0,98600	0,98821	0,99758	
60 sec	1,00243	0,99776	1,00412	1,00186	0,99957	
120 sec	0,99990	0,99724	0,96935	0,99019	0,99376	
300 sec	0,99614	1,00109	1,02919	1,01800	1,00250	
600 sec	0,99633	1,00471	1,01641	1,04040	0,98829	
	RMSE					
	RNTQ5 1(123)	RNTQ5 2(123)	RNTQ5 1(345)	RNTQ5 2(345)	RNTQ6 1(123)	RNTQ6 2(123)
10 sec	0,13667	0,12489	0,09857	0,09556	0,14834	0,13657
30 sec	0,23267	0,21486	0,16776	0,16409	0,25896	0,23495
60 sec	0,37787	0,33957	0,24624	0,23788	0,41802	0,37535
120 sec	0,48037	0,44531	0,34441	0,33561	0,53088	0,48929
300 sec	0,88202	0,77885	0,57589	0,56644	0,97724	0,84560
600 sec	1,52090	1,34731	0,84160	0,85820	1,68894	1,51960
	RNTQ6 1(456)	RNTQ6 2(456)	RNTQ7 1(1234)	RNTQ7 4(1234)	RNTQ7 1(4567)	
10 sec	0,09686	0,09518	0,14502	0,13454	0,09769	
30 sec	0,16530	0,16235	0,26180	0,23805	0,16958	
60 sec	0,24307	0,23569	0,41279	0,37276	0,24820	
120 sec	0,34258	0,33311	0,51314	0,48755	0,34360	
300 sec	0,57089	0,57332	0,99804	0,80704	0,59433	
600 sec	0,80621	0,82050	1,55210	1,36273	0,78315	

Table 4: RNTQ estimators performance at BM simulation with sparse sampling

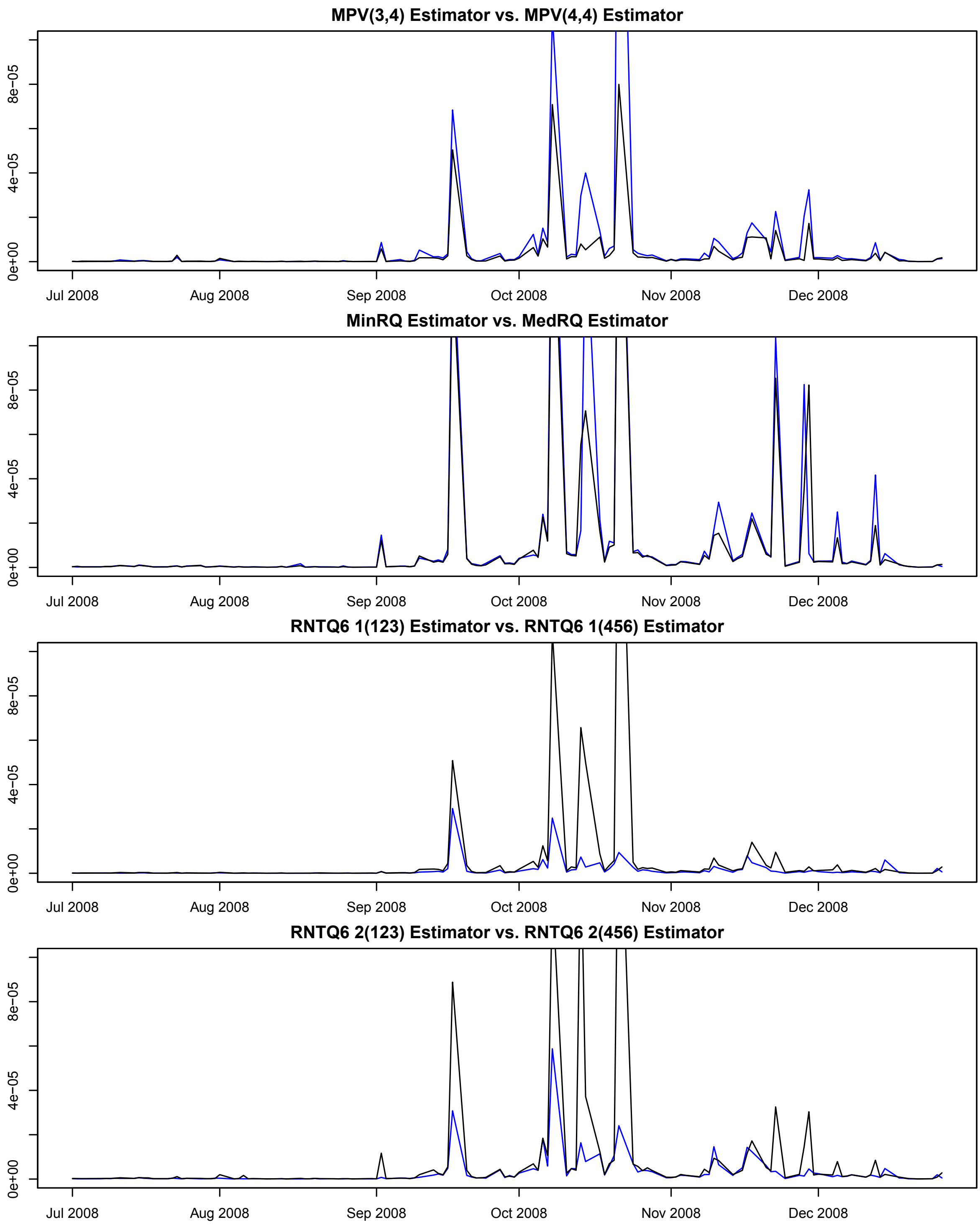

Figure 1: Integrated quarticity estimations for Pfizer during year 2008

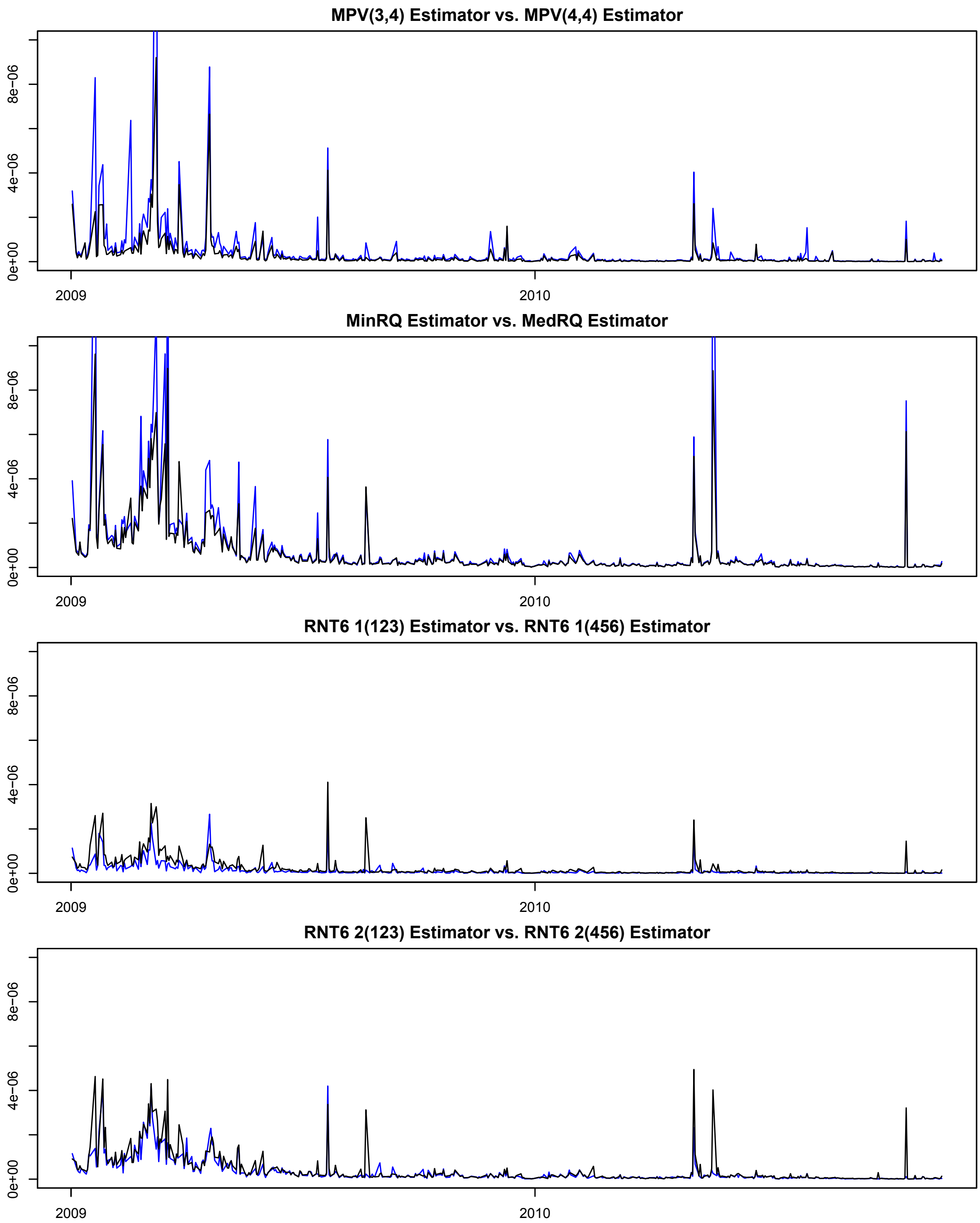

Figure 2: Integrated quarticity estimations for Pfizer during years 2009-2010

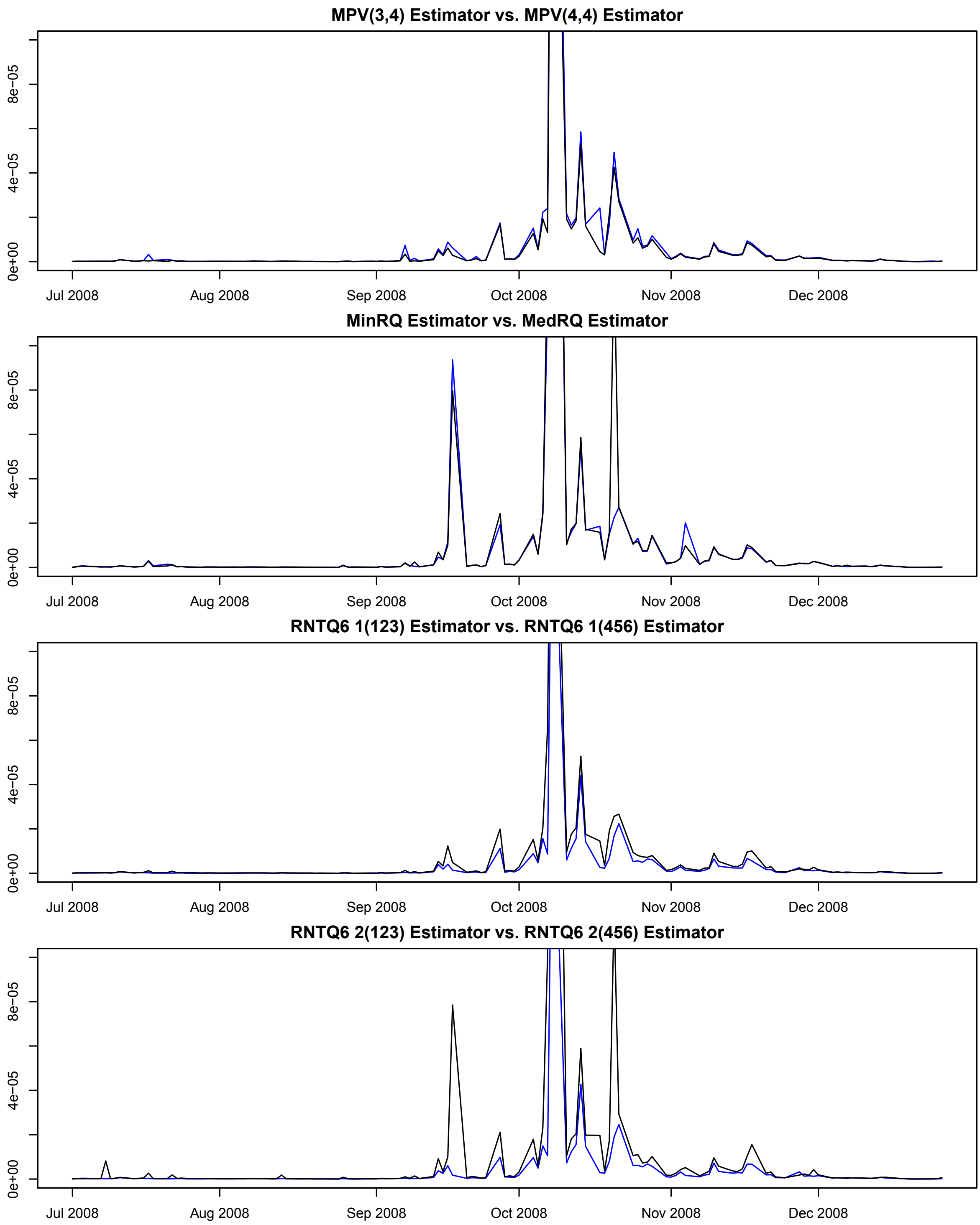

Figure 3: Integrated quarticity estimations for Exxon Mobil during year 2008

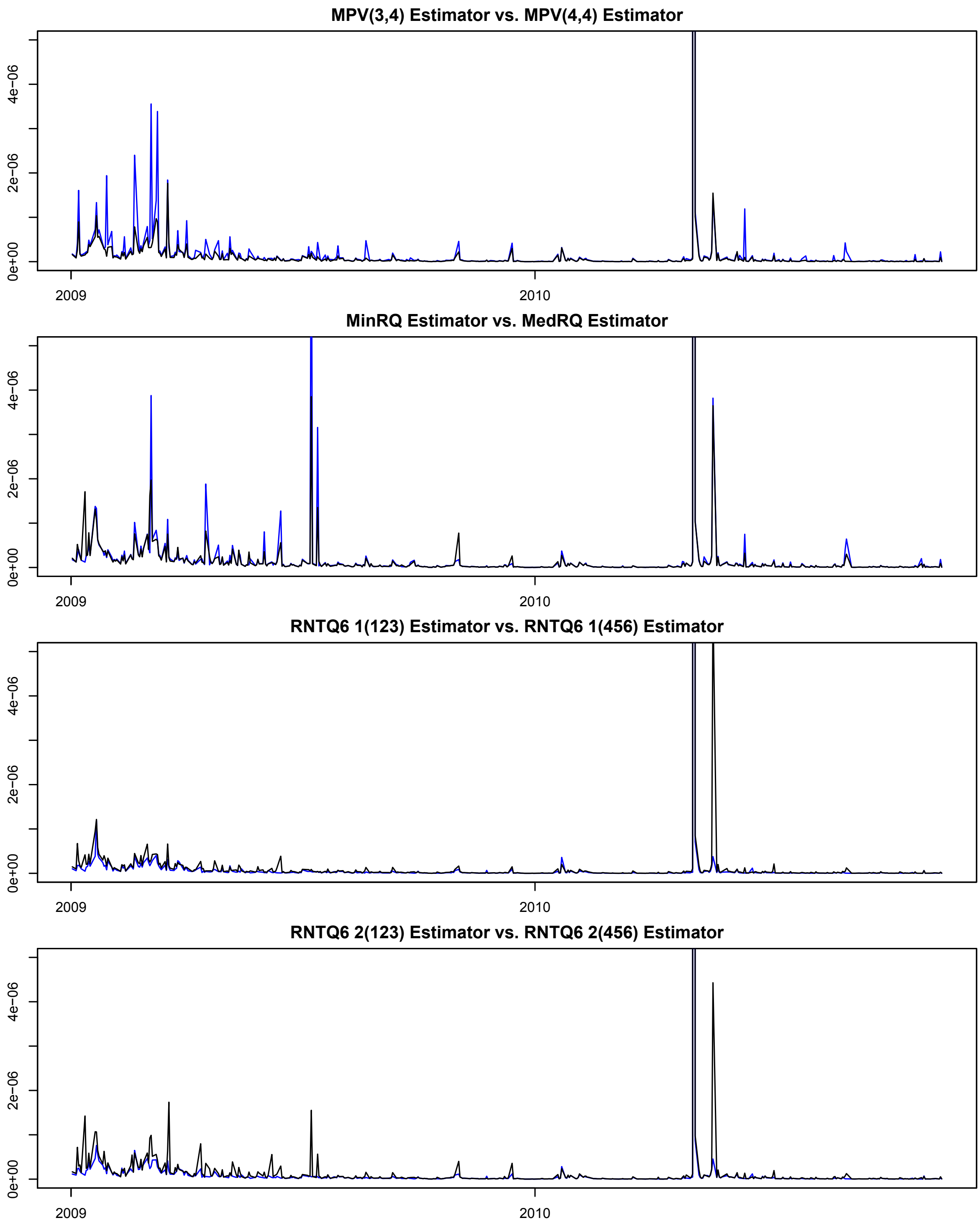

Figure 4: Integrated quarticity estimations for Exxon Mobil during years 2009-2010

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
1	**30,59377**	23,65482	10,32779	9,51000	32,83688	26,70508	9,18493	8,58909	31,22693	24,17492	10,69920
2	23,65482	**22,59096**	10,57821	9,88969	25,25456	24,24691	9,48675	8,95416	23,60849	22,35311	10,87703
3	10,32779	10,57821	**11,51716**	10,08218	10,01471	10,21933	11,03266	9,71237	9,97582	10,25680	11,01781
4	9,51000	9,88969	10,08218	**10,59576**	9,03519	9,29152	10,03514	10,33124	8,94469	9,44070	9,72751
5	32,83688	25,25456	10,01471	9,03519	**39,75629**	30,53404	8,77675	8,06995	38,36735	28,31851	10,43611
6	26,70508	24,24691	10,21933	9,29152	30,53404	**28,65126**	9,00023	8,30715	28,81516	26,62022	10,59904
7	9,18493	9,48675	11,03266	10,03514	8,77675	9,00023	**11,01310**	9,81208	8,76730	9,19042	10,69658
8	8,58909	8,95416	9,71237	10,33124	8,06995	8,30715	9,81208	**10,30176**	8,03246	8,53258	9,40506
9	31,22693	23,60849	9,97582	8,94469	38,36735	28,81516	8,76730	8,03246	**40,48261**	27,19568	10,39167
10	24,17492	22,35311	10,25680	9,44070	28,31851	26,62022	9,19042	8,53258	27,19568	**29,89296**	10,59005
11	10,69920	10,87703	11,01781	9,72751	10,43611	10,59904	10,69658	9,40506	10,39167	10,59005	**11,29826**

Notations:	RNTQ5 1(123)	1	RNTQ6 1(456)	7
	RNTQ5 2(123)	2	RNTQ6 2(456)	8
	RNTQ5 1(345)	3	RNTQ7 1(1234)	9
	RNTQ5 2(345)	4	RNTQ7 4(1234)	10
	RNTQ6 1(123)	5	RNTQ7 1(4567)	11
	RNTQ6 2(123)	6		

Table 5: Approximate values of asymptotic covariance matrix for some RNTQ5, RNTQ6 and RNTQ7 estimators applied to pure Brownian motion process

Bibliography

T. G. Andersen and T. Bollerslev. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, Nov 1998.

T. G. Andersen, T. Bollerslev, F. X. Diebold, and P. Labys. The distribution of realized exchange rate volatility. *Journal of American Statistical Association*, 96(453):42–55, March 2001.

T. G. Andersen, T. Bollerslev, and F. X. Diebold. Roughing it up: Including jump components in the measurement, modeling and forecasting of return volatility. *NBER Working Paper*, 2006.

T. G. Andersen, D. Dobrev, and E. Schaumburg. Jump-robust volatility estimation using nearest neighbor truncation. *NBER Working Paper No. 15533*, November 2009.

T. G. Andersen, D. Dobrev, and E. Schaumburg. Integrated quarticity estimation: Theory and practical implementation. May 2010.

T. G. Andersen, D. Dobrev, and E. Schaumburg. A functional filtering and neighborhood truncation approach to integrated quarticity estimation. *National Bureau of Economic Research Working Papers*, 2011.

G. Bakshi, D. Madan, and G. Panayotov. Deducing the implications of jump models for the structure of stock market crashes, rallies, jump arrival rates and extremes. May 2008.

F. M. Bandi and J. R. Russell. Microstructure noise, realized volatility, and optimal sampling. December 2003.

O. E. Barndorff-Nielsen and N. Shephard. How accurate is the asymptotic approximation to the distribution of realized variance? *Identification and Inference for Econometric Models*, June 2002.

O. E. Barndorff-Nielsen and N. Shephard. Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics*, Vol.2(No.1):1–37, 2004.

O. E. Barndorff-Nielsen and N. Shephard. Econometrics of testing for jumps in financial economics using bipower variation. *Journal of Financial Econometrics*, (4):1–30, 2006.

O. E. Barndorff-Nielsen, S. E. Graversen, J. Jacod, M. Podolskij, and N. Shephard. A central limit theorem for realised power and bipower variations of continuous semi-martingales. *Journal of Financial Econometrics*, Vol.2(No.1):1–37, 2006.

O. E. Barndorff-Nielsen, P. R. Hansen, A. Lunde, and N. Shephard. Realized kernels in practice: trades and quotes. *Econometrics Journal*, 12:C1–C32, 2009.

O. E. Barndorff-Nielsen, P. R. Hansen, A. Lunde, and N. Shephard. Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *Journal of Econometrics*, (162), January 2011.

J. A. Brogaard. High frequency trading and its impact on market quality. January 2012.

J. Cornelissen and K. Boudt. Rtaq: Tools for the analysis of trades and quotes in r. August 2010.

M. E. Drew, A. Marsden, and M. Veeraraghavan. Small firm effect, liquidity and security returns : Australian evidence. *Journal of Emerging Market Finance*, (5), October 2006.

M. Glezakos, K. Vafiadis, and J. Mylonakis. Analysis of intra-day volatility under economic crisis conditions. February 2011.

P. R. Hansen and A. Lunden. Realized variance and market microstructure noise. *JBES Invited Address*, August 2005.

H. Hong and J. Wang. Trading and returns under periodic market closures. *The Journal of Finance*, Vol.LV(No.1):297–354, February 2000.

J. Jacod, Y. Li, P. A. Mykland, M. Podolskij, and M. Vetter. Microstructure noise in the continuous case: The pre-averaging approach. *Stochastic Processes and their Applications*, (No.119):2249–2276, 2009.

B. G. Malkiel, A. Saha, and A. Grecu. The clustering of extreme movements:stock prices and the weather. *CEPS Working Paper*, (186), February 2009.

M. E. Mancino and S. Sanfelici. Estimation of quarticity with high frequency data. January 2012.

S. Mikhailov and U. Noegel. Heston's stochastic volatility model implementation, calibration and some extensions. *Wilmott Magazine*.

M. Podolskij and M. Vetter. Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. 2006.

P. Shi. Correcting finite sample biases in conventional estimates of power variation and jumps. 2009.

P. Tankov and E. Voltchkova. Jump-difussion models: a practitioner's guide.

T. A. Vuorenmaa. Decimalization, realized volatility, and market microstructure noise. May 2008.

L. Zhang, P. A. Mykland, and Y. Ait-Sahalia. A tale of two scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association*, 100(472), December 2005.

Selbständigkeitserklärung

Ich erkläre, dass ich die vorliegende Arbeit selbständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe.

Berlin, den 21.06.2012 Ivan Vasylchenko